Doctoral Dissertation

The Chinese urban economy and ASIE manufacture firms' activity: an empirical research based on the micro data

Graduate School of International Social Sciences
Yokohama National University

Baojun Fang

2021.12

Contents

Acknowledgement

All my achievement should be attributed to the love and friendship from my parents,friends, supervisors and Yokohama National University. Especially I am significantly motivated by the open and free academic atmosphere here which I never experienced before.

During the six years as PhD student in YNU, I am indebted to my supervisors, Professor Taro Akiyama, Masahito Kobayashi and Yoshiaki Omori. Especially I got many inspirations and encouragements from the tutoring of Professor Taro Akiyama, who support me to persist in my research into the topic which I really enjoy. Although I have experienced nearly 6-years long non-academic work in China, I regained my decisiveness and resilience to become a researcher in the future.

Graduate school of international social sciences of YNU is a prestigious research establishment in Japan, where I got many inspiring lectures. Besides activating and updating my stock of economics knowledge, I am enlightened in the field of statistics and econometrics that play the pivotal role in the empirical economic analysis. No one else ever thoroughly and clearly gave me the inspiration to dive into these two fields as in YNU.

Also I owe my new friends in Japan, who help me to better integrate into a different lingual and cultural circle. Without them I would not have overcome the anxiety and rootlessness in the first few years when I came here. Moreover outside the campus I deeply experience the differences between a open, inclusive and civilized society from an authoritarian society. Not only psychically people are also free to boast their own dreams and minds, which nourish the whole nation to advance in the development of wisdom and well-being.

Last but not the least, compared with the knowledge and skills, moral and integrity play more import roles in the human world, which are also the ethics source of economic thoughts. And I will uphold these virtues to contribute for the well-being of our world.

Acronyms

ASIE: refers to China's Annual Survey of Industrial Enterprises. As of July 2021, 1998-2013 period database is semi-overt for public.

CNKI: refers to China National Knowledge Infrastructure, a key national research and information publishing institution in China with the website as www.cnki.net.

EPO: refers to European Patent Office

NBS: refers to Chinese national bureau of statistics

HMT: refers to Hong Kong,Macau and Taiwan

LQ: refers to location quotient

CD: refers to cross-section dependence

CSMAR: refers to China Stock Market & Accounting Research

CODS: refers to China Organization Data Service, a affiliate to SMAR

SMAR: refers to State Administration of Market Regulation or 国家市场监管管理总局, a Chinese government agency

MAUP: refers to Modifiable Areal Unit Problem

SAC: refers to spatial autoregressive combined model

SDEM: refers to Spatial Durbin error model

Glossary

China/Chinese: refers to Chinese mainland not including Hong Kong,Macau and Taiwan.

State-owned: means that Chinese central/local governments full-fill the responsibility of the only shareholder/the largest percentage shareholder.

Collective-owned: means the entity that belongs to all the members of a group, usually a town or a village.The collective-owned firms are affiliated to a town/village government.

PATSTAT: refers to 2020 spring edition global patent database of version 5.15 by European Patent Office

TIANYANCHA: refers to 天眼查,a big data technology service company with a vast repository of Chinese enterprise information.

QICHACHA: refers to 企查查,a big data technology service company with a vast repository of Chinese enterprise information.
WANFANG: refers to 万方数据, an affiliate of the Chinese Ministry of Science & Technology, who provides access to a wide range of database resources.

Chapter 1

Introduction

Since 1990,due to deepening reform and opening up policy, Chinese enterprises have came through tremendous all-around changes during this process. Especially,due to the state-owned firms reform originating from 1992,the planning-economy style factories[1],were mostly shut down, merged, split or privatized.In addition, more and more HMT and foreign enterprises launched in Chinese market by sole investment or joined investment with Chinese peers. Accordingly, western corporate governance system and R&D system were gradually embedded in the economy plagued with Soviet Union economic system.From the Figure 1.1, we can have a glimpse about the overall evolution of the Chinese industrial enterprises during year 1998-2013.

In the second chapter,I will expand on the databases used in this paper, including ASIE, and patent databases of CNKI and PATSTAT. For any scholar interested in Chinese enterprise analysis, they have to face up to the problems with ASIE such as default value, clarifying blurred firms' identities, integrating scatted yearly databases as one, linking with other databases and handling abnormal values etc. Thanks to the data science, these obstacles have been mitigated and fixed somewhat. Meanwhile, using the applicant names, I have scratched the unique ID number assigned by government for nearly all the firm applicants in SIPO's patent database,which enable us to match the ASIE and CNKI.

Nevertheless, CNKI patent database only contains very limited information about the application record, where citation information is not publicized. To complement CNKI database, I resort to the PATSTAT by EPO(European Patent Office). PATSTAT consists of all-around information, however it is in different languages other than Chinese. Thereafter, I resort to Google Patent search engine to scratch all the Chinese names for the Chinese organisational applicant in PATSTAT. The linkage among these 3 database brings about the synergistic effect that more thorough analysis can be carried out.

[1]Those state-owned enterprises are literally factories,which focused on fulfilling the production quota arranged by the government instead of the market

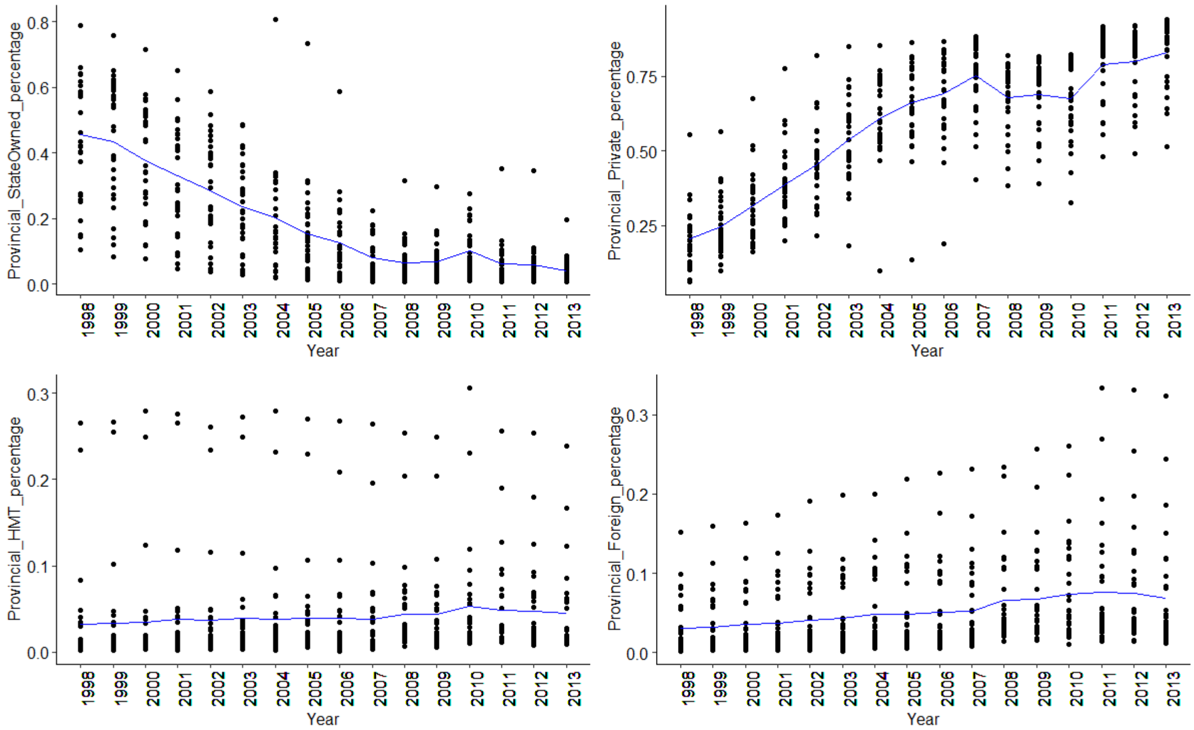


Figure 1.1: Ownership share transition

In the third chapter, I adopted two categories' indices respectively based on discrete space and continuous space pre-requisition to measure the agglomeration. Especially to embrace the location factor in a continuous space I scraped the geographic information for all the enlisted ZIP codes in ASIE, which makes the spatial econometric tools available in this paper.In addition,the relation between the firm size with the externalities are estimated in the city. And spatial weights matrices are constructed to introduce a spatial econometric analysis frame. Thereafter spatial spillover effect by Machiavellian and Jacobian externalities are estimated alongside with the basic panel models for the ASIE data. Especially different from basic models the spatial econometric models disentangle the direct effect from the indirect effect.

In the fourth chapter, co-patenting and citation networks are constructed based on the result of the second chapter. Moreover yearly descriptive statistics and visualization are achieved. Based on the derived results I will probe into the co-patenting and citation networks' influences on firms' performance. Especially the firm level Farrell production efficiency is estimated with adoption of a stochastic frontier model. In addition effects of each node's role is estimated with respect to efficiency, wage level and innovation performance.

Chapter 2

ASIE and the patent database

2.1 Introduction

The linkage of ASIE database with other database about Chinese industrial firms has always been a attractive but thorny domain for every researchers concerned with the empirical analysis about Chinese manufacturing sector. Due to the intricate composing process and shift of statistical system ASIE are also entangled with identification issues across different years. M, Eberhardt, et al.(2011) match both SIPO and USPTO patents of invent filed between 1985-2006 to a subset of about 20,000 firms in ASIE with the identification numbers recorded in Oriana, where a faction of ASIE firms are recorded with latin letter names, Chinese names and ids(i.e. id of ASIE). Other than the identifiers, J. Dang and K. Motohashi(2015) match SIPO's invent data with ASIE between 1985 to 2008 by means of the firms' names. Likewise enlarging the scope to 3 types of patent, Z. Xie and X. zhang(2014) match SIPO with ASIE between 1985 to 2009. Zl. He, et al.(2018) strengthened the method with fuzzy matching firms' names and increase the matched result for 3 types of patent between 1985 to 2009.

Except M. Eberhardt, et al.(2011) all other researchers focus on the firms' names to bridge the databases. Nevertheless modification of business registry info including names is very common in the long time span in China. So it is certain that the patenting fact must outnumber the matched results with names. Moreover each Chinese firm has its own unique identifier(i.e. organization code),which is also recorded in ASIE as id. With the social credit system developing firm's unique identifier can be easily accessed through the business registry query website like Tianyancha.

Meanwhile the non-Chinese firm names are the main obstacle for researchers to match the ASIE with patent database by EPO and USPTO. For those applications overseas filed by Chinese firms, it is very difficult to identify the Non-Chinese names(M, Eberhardt, et al.2011). And this issue can also be settled with the help of Google patent which align the patent application globally based on the patent family. Therefore it is very reasonable to trace the application overseas back to their domestic counterparts and snatch their Chinese names. In order to match these retrieved complementary data with the ASIE, web scraper and string fuzzy match algorithm are heavily utilized here.

Besides linking to ASIE, identifying the applicants unenrolled in ASIE are also achieved as the construction of co-patenting and citation networks are also target here, because both of these networks involve more applicants out of the reach by ASIE

2.1.1 ASIE database

According to NBS, ASIE's involves all the state-owned enterprises and the non-state-owned enterprises with turnover over 5 million Chinese yuan(20 millions Chinese yuan since 2011). And the industries can be divided into 3 subgroups of mining, manufacturing and public services(water, electricity and gas).[2] And by now the available database for public is based year 1998-2013.[3]

Due to the authority and continuity of this database's composition by NBS, ASIE is more frequently utilized to probe into the micro level issues with Chinese manufacturing sector compared with the database such as the small and medium sized enterprise database by CSMAR. As to the adoption of this database, the most mentioned problems can be summed up as bellow:

1. Non-continuous observations across different years. Some enterprises are missing halfway in this time span as a result of bankruptcy, merge & acquisition, public-owned firm's reform, slump of business and even missing response. Especially most research only used the database from 1998-2007, which can be attributed to the comparatively mature methodology developed by Loren Brandt, et al(2011) to integrate 1998-2007 databases. On contrary compared with the database before 2008, post-2008 database seemingly experienced drastic amendment of firm samples.

2. Inconsistency of the variables across different years just as showed by table 2.1. Among the these variables the ID variable is key to identify the unique firms. Nevertheless ID information is missing in year 2008. Also variables closely related this research like new product output(新产品产值), total intermediate input(中间投入) and payable wages(应付工资) are only available in some years. Specifically new product output are only

[2]The specific categorization were amended twice, which will be unified in accordance to GB/T 4754—2002 of NBS.

[3]This database is not totally exposed to public but is semi-exposed in some Chinese universities' internal database library such as Peking University.

available in year 1998-2007, 2009 and 2010; payable wages is missing in year 2010; total intermediate input is missing since 2008.

Table 2.1: Yearly counts of observations and variables

Year	1998	1999	2000	2001	2002	2003	2004	2005
Count of observations	165,116	162,033	162,885	171,256	181,557	196,222	279,092	271,835
Count of variables	105	105	106	103	95	88	150	140
Year	2006	2007	2008	2009	2010	2011	2012	2013
Count of observations	301,961	336,768	411,212	473487	348,536	303,392	310,832	344,882
Count of variables	130	140	72	72	80	98	88	97

3. Unstable variable's definition across the period. The most obvious one is the ID information, which refers to unique organization code(i.e. 组织机构代码) issued by NACAO affiliated to SAMR[4]. And to certify the code a certification rule was defined by official regulation GB 11714-1997 and another amended version in 2007. Except the occasion that registered content need to be modified, the code will stay unchanged for a unique legal person. So the recorded ID starting with AH, CH, BC, BD, BJ, BT,GX, etc are found to be unverifiable dummy IDs which cannot be linked to other years' entries. Moreover the definition of wage payable and welfare benefits payable were modified based on the amended version of accounting standard issued in year 2006. According to the newer version, wage payable and welfare benefits are combined as payroll payable alongside with social security expenses, education expenses, labor union fees, etc. Therefore payroll payable covers more than before.

4. Price deflation. As the responses by sample firms are based on values at the responding time point, the price level are not unified. Different from Loren Brandt(2011), I use the yearly composite PPI (i.e. producer price index for industrial products) publish by NBS with year 1998 as base level 100.

2.1.2 Patent database

The patent data involved here includes all the openly publicized invent applications documented at SIPO(i.e. China's State Intellectual Property Office) and overseas. They are all filed in by Chinese firms and universities alongside with other bibliometrically related applications by other types of organization. And the time span is set as the period with the ending date as December 31th 2013 according to the application date filed with.

[4]NACAO was reformed as CODS(China Organization Data Service) affiliated to SMAR(State Administration of Market Regulation or 国家市场监管管理总局)

Because the first patent protection law of China was promulgated in 1985, the first patent filed in can be traced back to 1985 and the starting date is set as January 1st 1985.

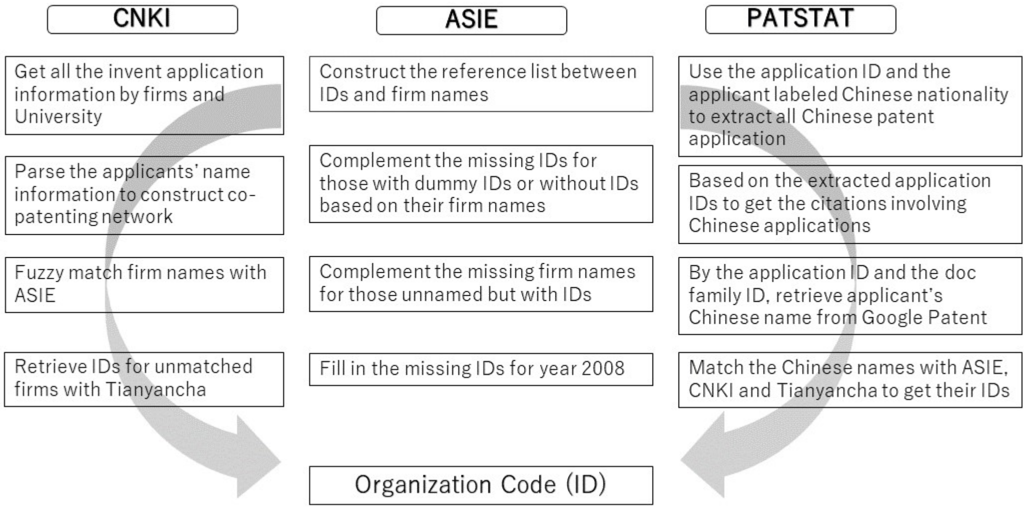
Data source here contains 3 openly accessible sources: search engine of CNKI, PATSTAT of EPO(i.e. European Patent Office) and search engine of Google patent. And they have their own obvious strengths and shortcoming as bellow:

Table 2.2

	Strengths	**Shortcomings**
CNKI Patent	1. It is downloadable in batch; 2. All applicant name are in Chinese which can be matched with ASIE more easily; 3. More frequently updated.	Only information of patent title, inventor name, applicant name, application data and publication date can be downloaded in batch.
PATSTAT	1.The citation relation is discreetly recorded among the publicized patent application globally; 2. All around information related to the application are listed.	1. Nearly all Chinese applicants' name are in non-Chinese messily and non-uniformly , which make it impossible to be linked with ASIE directly; 2. Nationality and type of applicant are not sorted up thoroughly.
Google Patent	1.Frequently updated all-round information are exposed altogether with the scanned original application document ; 2.Same family's patents are assembled across the different applicaton IDs, which make it convenient to trace back to the applicant names in its native language.	1. It cannot be downloaded in batch; 2. The information are scatterd and not sorted out.

2.2 Methodology

As the linkage is set up from scratch based on very tedious and maze-like process such as data wrangling, fuzzy match and visualization, it is very necessary to outline the process beforehand as figure 2.1 shows. All the operations I take is aimed to attach the organization IDs to all the firm applicants, which can be linked easily across different databases related to Chinese firms. Especially fuzzy match algorithm is a frequently utilized technique here to link firms across databases. It is introduced by Levenshtein(1965), which is the minimum number of single-character edits (insertions, deletions or substitutions) required to transform one word to another word. Therefore the Levenshtein distance between two strings a and b (of length $|a|$ and $|b|$



respectively) is given by $lev(a,b)$ where

$$lev(a,b) \equiv \begin{cases} |a| & if \quad |b| == 0, \\ |b| & if \quad |a| == 0, \\ lev(tail(a), tail(b)) & if \quad a[0] = b[0], \\ 1 + min(lev(tail(a), b), lev(a, tail(b)), lev(tail(a), tail(b))) & otherwise \end{cases} \tag{2.2.1}$$

And the tail of any string a is the substring without the first character of a, and $a[n]$ is the nth character of a counting from 0.So it needs $|a| \times |b|$ steps to get the Levenshtein distance between them. To clarify this obscure definition a practical example about two Chinese strings "核工业部北京核仪器" and "中国核工业北京核仪器" is listed as table 2.3 shows. The calculations starts form $[0, 0]$ to $[9, 10]$ in the matrix and the derived distance is 3, which implies these two names probably refer to the same establishment. This algorithm is especially sensible for Chinese misspelling and Chinese abbreviation such as "湖南省醴陵市太山出口鞭炮烟花制造有限公司", "醴陵市白兔潭大山花炮厂" and "醴陵市太山出口烟花制造有限公司" all refer to the same establishment. And "湖南省醴陵市白兔潭" is the location of this firm, which differs in each response alongside with "大" as a misspelling of "太". Inconsistent names can be attributed to formally modification for business registry or just abbreviation at the typewriter's will. To complement the Levenshein distance, names like "醴陵市白兔潭大山花炮厂" cannot be verified in open-accessible business registry query website like Tianyancha. So it can be concluded that this name is just a misspelling.

Table 2.3

	loc	0	1	2	3	4	5	6	7	8	9	10
loc	"" is null	""	中	国	核	工	业	北	京	核	仪	器
0	""	0	1	2	3	4	5	6	7	8	9	10
1	核	1	1	2	2	3	4	5	6	7	8	9
2	工	2	2	2	3	2	3	4	5	6	7	8
3	业	3	3	3	3	3	2	3	4	5	6	7
4	部	4	4	4	4	4	3	3	4	5	6	7
5	北	5	5	5	5	5	4	3	4	5	6	7
6	京	6	6	6	6	6	5	4	3	4	5	6
7	核	7	7	7	6	7	6	5	4	3	4	5
8	仪	8	8	8	7	7	7	6	5	4	3	4
9	器	9	9	9	8	8	8	7	6	5	4	3

2.3 ASIE raw data process

ASIE can be traced back to 1998, so it is very vulnerable to loose data entry, modification of business registry, amendment of statistical regulation and some subtle historical reasons. And the key clue to string these sparse observations is to attach the verifiable organization code.

Based on existing data entries, one verifiable organization code can be linked to one or multiple names but not vice versa. So for those entries only containing the IDs, their firm names can be easily retrieved from other years' proper and intact entries. And the unmatched IDs can be queried in Tianyancha, Qichacha, etc. As to the firm names without IDs, besides merging with the intact entries and business registry queried result, fuzzy match alongside with comparison of address, representative name, industry sector code and telephone number is heavily used because the misspelling and abbreviation seriously flaw the firms' identity.

Before the fuzzy match, a preliminary trim of firm names is needed as they are just auxiliary to identify a establishment. And these auxiliary information or substrings can be categorised into punctuation[5], administrative division level, registry types, ownership, which can be exemplified as bellow:

And also replicates is another obstacle encountered. The observations with no entry into variables of interest are deleted. For those with different address, they are treated as different plants belonging to the same firms. Trough this process a unbalanced panel data with organization codes as unique identifiers can be achieved

[5]It needs to be stressed unlike Latin letter expression space is of no importance in Chinese writing system.

Table 2.4

administrative division level	省，市，区，县，乡，自治
registry info	厂，公司，总公司，分公司，集团 股份有限公司，有限责任公司，有限公司，集团
ownership	国营，集体，私营，中外合资，中日合资，中美合资

with the yearly counts as tables 2.4 and 2.5 show. Especially extracted from these $4,224,492$ observations $796,833$ firms have been enlisted at least once in ASIE. And $124,751$ firms are just enlisted once alongside with $20,064$ firms outliving these 16 years.

Table 2.5: Yearly count of observations

year	Freq.	year	Freq.
1998	164,947	2006	301,851
1999	161,946	2007	336,670
2000	162,800	2008	411,823
2001	168,946	2009	320,522
2002	181,481	2010	310,581
2003	196,154	2011	302,582
2004	276,283	2012	311,308
2005	271,772	2013	344,826
Total	**4,224,492**		

Table 2.6: count of firms grouped by times of occurrence

#	Freq	Percent	#	Freq	Percent
1	124,751	16	9	34,486	4.33
2	120,444	15	10	35,850	4.5
3	110,024	14	11	21,409	2.69
4	60,314	7.57	12	16,930	2.12
5	73,467	9.22	13	18,795	2.36
6	48,628	6.1	14	7,863	0.99
7	55,065	6.91	15	10,007	1.26
8	38,736	4.86	16	20,064	2.52
Total	**796,833**				

2.4 Construction of co-patenting network

2.4.1 Linkage of ASIE and CNKI patent

Data source here is based on the published 1985-2013 invent patent by SIPO through the CNKI patent platform. And all the applications with applicant names containing "公司"(i.e. firm), "厂"（i.e. factory） and "大学"（i.e. university) are retrieved. Compared with the published database by SIPO, the CNKI patent is a online query database. And it can be verified that the records are more frequently updated in CNKI, which means the newly publish ones can be included in time. By fuzzy matching firm names with ASIE and query into the business registry database, organization codes can be attached mostly to the firm applicants of the CNKI's data.

Another difficulty entangled is to attach the IDs with the firms names only used in 1980s and early 1990s. As Chinese state-owned and collective-owned have gone through dramatic reform and restructure in 1990s, overwhelming modification of business registry handicap the effort to retrieve the IDs for these firms. To address this issue, organization database of WANFANG[6] is introduced here as it hold plentiful organization evolution info. So the former names recorded in CNKI can be connected to their existing heirs according to WANFANG. The remained few unmatched ones can be either linked manually or just labelled with pseudo identifiers. For example "江汉石油管理局总机厂" is the predecessor of "中石化石油机械股份有限公司江汉机械厂", so these two names are taken as the same firm.

Compared with other researchers' matched results, a overview can be listed as bellow. Moreover a specific break down can be referred to table 3.2.

Table 2.7: Compared with existing match results[7]

	Period		Matched Result		
Authors	**ASIE**	**SIPO**	**# of ASIE Assignee**	**Identified # of Invent**	**Matched # of Invent**
Eberhardt et al.	1999-2006	1985-2006	1,219	405,180	44,344
Dang & Motohashi	1998-2008	1998-2008	12,208	283,377	126,386
Xie & Zhang	1998-2009	1985-2009	11,631	265,713	127,542
He et al.	1998-2009	1998-2009	30,711	536,956	257,032
My result	1998-2013	1985-2013	70,420	1,696,818	797,424

2.4.2 Co-patenting network

Based on the result of linking ASIE to CNKI, co-patenting refers to the application by over one applicants, which can be distinguished by applicants names that contain semicolons （i.e. ;）. So applications like CN201010563877 applied by "河海大学; 江苏省交通规划设计院有限公司; 南京河海科技有限公司" are co-patents. And there is a co-patenting network as figure 4.3 shows:

With parsing all applicants' names, application containing semicolons are all divided into applications that

[6]万方数据, an affiliate of the Chinese Ministry of Science & Technology

[7]The matched number of invent and the matched number of ASIE firms are calculated based on the percentage by Xie & Zhang; the matched number of invent is adjusted by the applicant number published by He et al. Due to absence of application number, I use patent names, applicant names, application date and publish date to drop the

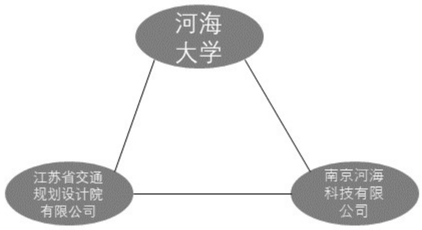


Figure 2.1

one application ID attaches to one applicant name. And it still needs discretion with the the parsed result. For the case of application ID CN93119522.5 that the applicant name is "广西中医学院; 龙胜各族自治县轻工业局猕猴桃制品厂; 苏必忠", one parsed result is a person's name "苏必忠". And in the case of application ID CN200910025687.2 with the applicant name recorded as "江苏省交通规划设计院有限公江苏省长江公路大桥建设指挥部; 中国人民解放军理工大学", one extremely lengthy parsed result is "江苏省交通规划设计院有限公江苏省长江公路大桥建设指挥部". And it can be inferred that it is a typo of "江苏省交通规划设计院有限公司; 江苏省长江公路大桥建设指挥部". As most Chinese names' lengths are less than 4, individual applicants can be parsed out easily. As of the typos, all the string length outliers are checked manually.

After the data wrangling and parsing, fuzzy matching with ASIE and inquiry into Tianyancha are carried out. Except those firms only existing in 1980s and early 1990s, nearly all firms are attached with their IDs. Thereafter the edge lists with IDs are sorted out. A brief overview of the linkage result is just as table 6.1 and table 6.2 show in the appendix. And among the total count of 914,591 invention patents, 782,227 can be matched with ASIE enlisted firms which mounts to 70,420. Within the constructed co-patenting network, 6,910 out of 22,795 net nodes are ASIE enlisted firms.

2.5 Construction of citation network

2.5.1 Linkage of ASIE and PASTAT patent

Data source here contains PATSTAT Version 5.15 and Google patent.And the linkage focus on the patent applications recorded in the citation documents. Moreover the linked citation records here is centering on the Chinese firms, which means that at least one node of a citation link is a Chinese organization and neither node is individual. Furthermore invent, design and utility are taken in consideration, as citation between different types of patent is not uncommon. According to the catalogue of EPO citations have different origins, and only

duplicate records.

citations that are introduced by applicants with publicized verifiable application IDs are adopted here. Those citations with unique citing and cited $DOCDB\_Family\_IDs$ are taken as one effective citing linkage, as all the applications labelled with same $DOCDB\_Family\_ID$ refer to the same technology content.

The most prohibitive matter with adopting PATSTAT's Chinese applicatons is that nearly all the Chinese applicants' name are in non-Chinese languages. Moreover same firm's name may varies dramatically across different applications such as *"HUWAEI TECHNOLOGIES CO., LTD."*, *"KHUAVEJ TEKNOLODZHIZ KO., LTD."* and *"HUAWEI TECHN CO., LTD."*. Actually all these names refers to *Huawei Technologies Co., Ltd*(i.e. 华为技术有限公司). Nevertheless *"huawei techniques co., ltd."* refers to a Taiwan's firm "华威科技股份有限公司". Obviously the non-Chinese translations can not keep all the nominal information such as tones. Therefore it is quite unreliable to adopt the back-to-Chinese translation to identify the firms.

Secondly EPO doesn't label the nationality for all the applicants. And those applicants with applicants filled in China National Intellectual Property Administration(i.e. SIPO) have high propensity to be Chinese. Therefore retrieving the original Chinese names recorded in the documented files is the most convincing and reliable way to identify them. Meanwhile Google patent search engine is open-accessible and all-around maintained which make it possible to retrieve the Chinese names.

Thirdly, identifying non-Chinese applicants is another issue, although EPO has made some but not enough efforts. For example, *"via technologies, inc."*,*"weisheng electrionic co., ltd."* and *"vasion electronics co., ltd."* are taken as different companies in PATSTAT. Nevertheless,they refer to the same firm *"VIA Technologies"* in Taiwan. Similar mistaking can also be exemplified by *"sumitomo electric k.k."* and *"sumitomo electronic industries, ltd."*.

To solve the mentioned three defects, I resort to the Google patent for Chinese character names of applicants with nationality of China and Taiwan as well as English names of other countries' applicants. Then as what is done with CNKI, the organization codes are attached. With regards to other countries' applicants, their English names are used to identify same firm across different applications alongside with the PATSTAT Standardised Names. Even though fuzzy match can detect misspelling and abbreviation, it cannot be assured that all the identification can be soundly finished. For the linkage between likes of *"Orange SA"* and *"France Telecom S.A."*, fuzzy match is totally powerless. Because France Telecom S.A. initiated to rebrand them as Orange S.A.

After closely manual check with non-Chinese applicants, a sketch of citation networks centering Chinese firms surface here. Among these citation links same-directed links between same applicants as well as same $DOCDB\_Family\_IDs$ are abundant and taken as one effective link. Consequently it is very easy and pleasant to link the whole network to ASIE through the organization codes. And there are 316,785 patents embedded in this network out of which 115,456 are applied by 17,684 Chinese firms. Moreover 8,382 are ASIE enlisted firms.The specific result of nationalities of applicants are listed as tables 3.3, 3.4, 3.5 and 3.6.

As to the citations between different countries, tables[8] bellow give a ranking result about the nationalities of applications citing Chinese ones and the nationalities of those cited by Chinese ones respectively. Obviously US, Japan, Taiwan, South Korea and Germany maintain the most active interactions with China. A detailed breakdown of yearly nationalities shift can be traced in the tables of 6.3, 6.4, 6.5 and 6.6.

Citing Chinese patents

Rank	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	US	US	US	US	US	US	US	US	US	US	US	US	US	US	US	US
2	JP	JP	JP	JP	JP	JP	TW	TW	TW	TW	TW	TW	JP	JP	JP	JP
3	DE	CH	CA	CA	KR	TW	JP	JP	JP	KR	JP	JP	TW	TW	TW	KR
4	HK	DE	DE	KR	TW	KR	KR	KR	KR	JP	KR	KR	KR	KR	KR	TW
5	CA	GB	GB	CH	DE	DE	FR	DE	DE	DE	DE	DE	DE	DE	DE	DE

Cited by Chinese Patents

Rank	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	US	US	US	US	US	US	US	US	US	US	US	US	US	US	US	US
2	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP	JP
3	DE	CA	DE	DE	DE	TW	TW	TW	TW	TW	TW	TW	TW	TW	TW	TW
4	FR	DE	GB	CA	SE	KR	KR	KR	KR	KR	KR	KR	KR	KR	KR	KR
5	GB	KR	KR	SE	KR	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE

2.6 Summary

With the adoption of techniques from data science and taking the organization codes as the unique identifiers longitudinally across different databases, an enlarged linkage between ASIE and SIPO's database is achieved for a longer time span. Compared with the existing linkage result, all the data sources can be checked online. Moreover the ASIE database from 1998-2013 are more integrated by the unique organization codes and the linked results are prolonged to year 2013 which covering the early stage of patent application outburst since 2009, which causes at least doubled matched records.

Although patent citation analysis has been gaining considerable traction over the past few decades, micro level analysis about China are mostly limited to the context of patents themselves. There is no open published data bridging with other micro-level databases. The matched result here can fill the blank somewhat. Especially Chinese organization's identifying issue across different lingual contexts is somewhat tamped. The linked result can give a new insight into the technical progress rooted in firm's performance.

[8]country codes are in accordance with ISO 3166 code

Chapter 3

Manufacture sector geographic distribution and its influence on firm size in China

3.1 Introduction

Not only Geographic scientist, but also economists are attracted to the geographic distribution,especially the agglomeration phenomenon.Nevertheless,researchers have reach a deterministic conclusion to measure the extent of the agglomeration,which can be observed in different perspective.

Duranton and Overman(2002) argued that a satisfactory spatial agglomeration measurement should have the following properties:(1)be comparable across industries,(2)control for overall agglomeration trends across industries,(3)separate spatial concentration from industrial concentration,(4)be unbiased with respect to the degree of spatial aggregation,(5)admit a clear statistical significance test.

Combes et al. (2008, Chapter 10) and Kominers (2007) provide three other three additional properties:(1)be unbiased with respect to arbitrary changes to industrial classification;(2) be computable in the closed form from accessible data, and (3) be justified by a suitable model.

As far as now,agglomeration measures can be sub-grouped into 2 categories:model-based measurement and axiomatic-approach-based measurement. The first group rely on the assumption that each location is discrete and equidistant spatial units. The second group take the distance between each firms into consideration, with taking firms' location in a continuous space. Obviously the second group are independent of the administrative boundaries and closer to the reality, but much more demanding on the data and very hard to grasp mathematically.In the next section I will give a introduction about these two groups' indices utilized in this paper.

Cecile Gaubert(2018) establish a equilibrium model of heterogeneous firms that freely mobile within a country and can choose the size of the city where they produce. Especially the labour force size in a given city can bring in positive externalities and mounting pressure for raising wages as the land supply is limited. Therefore firms sort across cities of different sizes in order to maximise their profits. In this process city labour size's relation with the firm's employee scale and profit is modelled.

Kim (1995) and Holmes and Stevens (2002) find a positive correlation between industrial agglomeration and plant size both across and within manufacturing industries by using plant-level data in the United States.

Dongya Li, et al.(2011) identified that industrial agglomeration has a positive and statistically significant causal impact on firm size through the instrumental variable estimation based on the ASIE data from 1998-2005.

Based on the discrete agglomeration results, a spatial econometric frame is adopted here to probe into the spillover effect about the agglomeration on firm size among the neighbouring cities.

3.2 Methodology

Location Quotient indices(LQ indices)

The discrete agglomeration indices are very charming for policy-makers, who are keen on making regional policies for districts, cities, provinces or countries. Nevertheless,these indices are clumsy with handling the data within a specific space unit. In addition, all discrete agglomeration indices fail to address the Modifiable Areal Unit Problem(MAUP). The MAUP make an a discrete agglomeration index awkward to cope with the agglomeration phenomenon across different administrative regions.

The location quotient can be used to see whether the employment of a specific industry is above or bellow the average overall. The LQ offers the opportunity to compare industries which differ in size because of the calculation using the shares on the magnitudes (Vgl. FIGUEIREDO, O.; GUIMARAES, P; WOODWARD, D. 2007). The LQ for the industry A in subregion i can be calculated bellow, where the superscript N bellow means the whole region.

$$LQ_i^A = (E_{i,A}/E_i)\Big/(E_{N,A}/E_N) \tag{3.2.1}$$

And $E_{i,A}$=Industry A's employment in subregion i, E_i=Overall employment in subregion i, $E_{N,A}$=Industry A's employment nationally, E_N=Overall employment nationally.
The drawbacks of the LQ indices are also obvious:1.LQ arbitrarily take unit one as the cut-off value to determine whether a agglomeration exits in a region, which is always questionable because this industry probably does not stick out at all compared with other industries. 2. LQ cannot solve the MAUP problem.3.LQ does not take the firms' size into consideration,a single firm's extremely big size can seriously affect the meaning of the LQ indice. To cope with these drawbacks,several enhancements have been applied to the LQ.

Firstly,Moineddin,Beyene and Boyle (2003) suggested a testing method for significance of LQ's result,with the assumption that: (1) in a subregion i, each employee's choice about whether to work for industry A or not follows a Bernoulli distribution with parameter $p_{i,A}$ independently; (2) for a subregion i beholding E_i's workers, the outcome number of being engaged in industry A is in a binomial distribution(i.e. it is asymptotically a normal distribution, as E_i, $E_i \times p_{i,A}$ and $E_i \times (1 - p_{i,A})$ are large enough;(3) the distribution of each subregion is independent. Then by means of delta method, the variance of LQ_i^A can be estimated. The variables involved are defined as bellow:

$$\begin{aligned} g_{i,A} &= E_{i,A}/E_i \\ g_A &= E_{N,A}/E_N \\ LQ_i^A &= g_i^A/g^A \\ E(g_i^A) &= p_{i,A} \\ E(g_A) &= p_A \end{aligned}$$

Consequently,based on the Tyler series of LQ_i^A, the variance can be approximated as bellow:

$$\begin{aligned} var(LQ_i^A) \equiv var(g_{i,A})/g_A) &\approx \frac{var(g_{i,A})}{{p_A}^2} + \frac{p_{i,A}^2 var(g_A)}{{p_A}^4} - \frac{2p_{i,A}cov(g_{i,A}, g_A)}{E_N \times p_A^3} \\ &= \frac{p_{i,A}(1 - p_{i,A})}{E_i \times p_A^2} + \frac{p_{i,A}^2 \sum_{j=1}^{n} E_j p_{j,A}(1 - p_{j,A})}{E_N^2 \times p_A^4} - \frac{2p_{i,A}^2(1 - p_{i,A})}{E_N \times p_A^3} \end{aligned} \tag{3.2.2}$$

Thereafter, the (1-α) confidence interval for the LQ_r is $g_r/g \pm Z_{\alpha/2}\sqrt{v(LQ_r)}$, and $Z_{\alpha/2}$ is the $\alpha/2$ percentile of standard normal distribution.

O'Donoghue and Gleave(2004) contributed the standard location quotient to contain the same issue, with the statistic as bellow:

$$SLQ_{ir} = (LQ_{ir} - \overline{LQ_i}) \Big/ std(LQ_i) \tag{3.2.3}$$

$$SLLQ_{ir} = (\log(LQ_{ir}) - \overline{\log(LQ_i)}) \Big/ std(\log(LQ_i)) \tag{3.2.4}$$

If LQ is distributed normally, the SLQ is a z-statistic. Otherwise, the logarithmic of LQ quotient is used to be transformed to SLLQ. In any case, the data must be verified by the normality test.The shortcoming is very obvious that the not all industries' distribution can be normal.Due to the heterogeneity of different industries, this assumption is strongly challenged. To overcome this, the bootstrap method is used here to contain this issue, as there is no any pre-assumption about the the distribution of SLQ or SLLQ. Especially,the the mean of 95 percentile cut-off value and its estimate standard error are derived here to construct a confidence interval for the 95 percentile cut-off value. Therefore, the subregion with the quotient outnumbering the cut-off value is reasonably to be taken as the industry i agglomeration region.

The enhancements above don't solve the firm size problem which is close to the level of market competition. Holmes and Stevens (2002) introduced a method for decomposition of the LQ to overcome that obstacle. The LQ is decomposed to its components and put together in a new way. Besides the employment number of different industries in different subregions, the count of firms and their employee's count are required. The composition of the LQ is as bellow:

$$Q^x_{ir} = Q^n_{i,r} \times Q^s_{i,r} \tag{3.2.5}$$

where

$$\begin{aligned}
Q_{ir} &= LQ_{ir} \\
Q^n_{i,r} &= (n_{ir}/E_r) \Big/ (n_i/E_n) \\
Q^s_{i,r} &= (E_{ir}/n_{ir}) \Big/ (E_i/n_i)
\end{aligned}$$

Among the variables here, newly added includes: Q_{ir} denotes the employment location quotient of industry i in subregion r, n_{ir} denotes the number of industry i's firms in subregion r, n_i denotes the the number of industry i's firms in the total region. Consequently through this process,the LQ quotient is decomposed to two sources: (1) $Q^n_{i,r}$ the ratio that the number of industry i's firms per capita divided by the total region's level; (2) $Q^s_{i,r}$the ratio that the average size of industry i's firms divided by the total region's level.Then take the natural logarithm of function 3.2.5 and let the lower case q's represent their counterparts, the function below can be derived:

$$q^x_{ir} = q^n_{i,r} + q^s_{i,r} \tag{3.2.6}$$

Then they use the β-coefficients of regressingq^n_{ir} on q^x_{ir} alongside with those of regressingq^s_{ir} on q^x_{ir}, which are as below:

$$\beta^n = cov(q^x_{ir}, q^n_{ir}) \Big/ var(q^x_{ir}) \tag{3.2.7}$$

$$\beta^s = cov(q^x_{ir}, q^s_{ir}) \Big/ var(q^x_{ir}) \tag{3.2.8}$$

Due to the function 3.2.6, $\beta^n + \beta^s = 1$. Especially, if ether of these two $\beta' s$ equals to zero, then the variance of the stand LQ quotient is totally accounted for by the other component. Moreover, β^s the relation between the agglomeration and firm's average size gets more attention. Furthermore, other than the average size of firms,a firm level localization is introduced to estimate the effect of subregion's localization quotient $q^x_e \equiv q^x_{ir}$

on a individual firm size ,where e refers to firm e as well as industry i and location r of this firm. Accordingly for any firm with size E_e, the firm level size quotient and the β-coefficient are defined as bellow:

$$q_e^s = ln[E_e \Big/ (E_i/n_i)] \tag{3.2.9}$$

$$\beta_e^s = cov(q_e^x, q_e^s) \Big/ var(q_e^x) \tag{3.2.10}$$

Nevertheless, even if the distribution of firm size is independent from the location of industry, under the case that there are just a infinite number of firms,the β_e^s and β^s is biased due to big-sized firms. In order to contain this issue, a *excluded localization quotient* $\tilde{Q}_e^x$ is introduced that current firm is excluded when calculating the standard LQ. Meanwhile, *a excluded firm-size level quotient* $\tilde{Q}_e^s$ is defined as well. And the natural logarithm of these two quotients are as bellow:

$$\tilde{q}_e^x = ln[(\frac{E_{ir} - E_e}{E_r - E_e}) \Big/ (\frac{E_i - E_e}{E_n - E_e})] \tag{3.2.11}$$

$$\tilde{q}_e^s = ln[E_e \Big/ (\frac{E_i - E_e}{n_i - 1})] \tag{3.2.12}$$

Therefore, with regressing $\tilde{q}_e^s$ on $\tilde{q}_e^x$,the coefficient $\tilde{\beta}_e^s$ can be utilized to find the correlation between a firm's size with its decision whether to locate a subregion with high LQ value. Moreover just as q_{ir}^x, $\tilde{q}_e^x$ can also be decomposed to two components as below:

$$\tilde{q}_e^x = \bar{q}_e^n \times \bar{q}_e^s \tag{3.2.13}$$

where

$$\bar{q}_e^n = ln[(\frac{n_{ir} - 1}{E_r - E_e}) \Big/ (\frac{n_i - 1}{E_n - E_e})]$$
$$\bar{q}_e^s = ln[(\frac{E_{ir} - E_e}{n_{ir} - 1}) \Big/ (\frac{E_i - Ee}{n_i - 1})]$$

Therefore, with regressing $\tilde{q}_e^s$ on ether of the components above, we can estimate that which component accounts more for the decision making of current firm. For a firm e, it may choose a low LQ valued subregion to gain potential market share or to a high LQ valued subregion to gain the benefit brought out by the agglomeration there, which may be mainly accounted for by a large number of peers there or just a small number of large-scaled firms.

Continuous-space localization indices

Based on the kernel density estimation and simulation, Gilles Duranton and Henry G. Overman(2004) introduce a new method to estimate the localization on a continuous space. Especially, they make some enhancements over discrete indices that (1) no evenness presumption for each subregion which means that localization caused by area, population size, natural resources or specific industry characteristics are not accounted for here;(2) Each firm's location is handled independently and not aggregated in some boxes(provinces, cities, counties, etc.),so the Modifiable Area Unit Problem is ruled out ;(3) The overall industrial localization tendency is compromised as the evolution of a industry's overall location pattern can be updated periodically in its calculation process. (4) Without any distribution presumption, Monte Carlo approach is utilized to simulate the pattern of randomness for each industry, therefore the non-randomness(i.e industrial localization) can be measured with the statistical significance of departure. Nevertheless, this is method is very demanding on the geographic information about each firm enrolled.

The first step is to construct kernel estimates of k-densities centering on bilateral distance between each pair of n's firms within current industry. Especially, the distance is a proxy for the true approximating time and pecuniary cost between them, so it may be distorted by the distance algorithm used[9] and the actual traffic conditions such as topography and road density. And the estimator can be listed as bellow:

$$\hat{K}(d) = \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} f(\frac{d-d_{i,j}}{h}) \tag{3.2.14}$$

$d_{i,j}$ is the distance between firm i and j, h is the chosen bandwidth and f is the kernel function. As the distance is absolutely above zero, in order not to underestimate the density around the bounder zero they adopt the reflection method by Silverman (1986).

Secondly, although function (3.2.14) is a U-statics like shaped and firms are presumed to locate independently, the bilateral distance are still dependent.Therefore, they adopt the Monte Carlo simulation to measure the departure from the randomness.

Thirdly, other than re-sampling the original distance calculated from census data, it is simulated that each firm randomly draws the existing sites of the industry they belong to without replacement as one existing location can only hold a firm in reality.Thereafter the bilateral distance can be calculated.

As to the construction of confidence interval, it has local confidence interval and global confidence interval constructed for the $d \in [0, \tilde{d}]$, where $\tilde{d}$ is the median among all the bilateral distances. Local confidence interval is to take the $\alpha$and 1-αpercentile with ascending the density of the distance d for all the simulations simultaneously. And the lower bound and upper bound are denoted as $\bar{K}(d)]$ and $\underline{K}(d)$. So it is reasonable to use these outcomes as counterfactual to be compared with the real density estimated by the function (3.2.14). Nevertheless, due to the autocorrelation between distances and the smoothing techniques,the agglomeration is correlated with the dispersion which is against the randomness, other than take the percentile on each distance, global confidence interval is to construct the upper and lower bound that only αsimulations can hit the upper bounds or lower bound, which are denoted as $\bar{\bar{K}}(d)]$ and $\underline{\underline{K}}(d)]$in the interval of $[0, \tilde{d}]$. Accordingly for industry i,the indices of localization is defined as bellow:

For local confidence interval:

[9]Actually in their paper, they calculate the distance based on a euclidean space projected by the curvature of UK on the earth, which could cause the systematic error.

localization index

$$\gamma(d) \equiv \max(\hat{K}(d) - \bar{K}(d), 0) \tag{3.2.15}$$

as well as dispersion index:

$$\psi(d) \equiv \max(\underline{K}(d) - \hat{K}(d), 0) \tag{3.2.16}$$

For global confidence interval:

localization index:

$$\Gamma(d) \equiv \max(\hat{K}(d) - \bar{\bar{K}}(d), 0) \tag{3.2.17}$$

dispersion index:

$$\Psi(d) \equiv \begin{cases} \max(\underline{\underline{K}}(d) - \hat{K}(d), 0) & if \sum_0^{\tilde{d}} \Gamma(d) = 0, \\ 0 & otherwise \end{cases} \tag{3.2.18}$$

Where we can notice that if a industry's $\hat{K}(d)$ surpass the upper bound anywhere in the interval $[0, \tilde{d}]$,dispersion is not accounted for with respect to this industry. To contain the employment into the density estimation for industry i, the density can be defines as bellow:

$$\hat{K}_{emp}(d) = \frac{1}{h \sum_{i=1}^{n-1} \sum_{j=1+1}^{n} e(i)e(j)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} e(i)e(j)f(\frac{d - d_{i,j}}{h}) \tag{3.2.19}$$

where e(i) and e(j) are the numbers of employee owned by firm i and j. To sum all the localization and dispersion index respectively across the distance d within the range of $(0 - Med)$, we can get the cross-distance localization and dispersion indices for a given industry sector.And Med is the median distance among all the establishment. Likely with summing all industries' localization and dispersion indices for a given distance level d in the same the range above, the composite localization and dispersion indices for distance level d can be derived.

3.2.1 Test of spatial autocorrelation

Traditional regression models mostly stipulate the assumptions about exogenousity of regressors, scedasticity and serial correlation restriction on the residual part. Whereas the cross-section dependence (abbre. CD) issue is absent in these literatures. Nevertheless our purpose here is just to probe into the cross-section relation(i.e. spatial autocorrelation)as the sections here are cities. The test used to detect cross-section dependence can be categorized into 2 groups:

1. cross-section data for spatial observation
For a *priori* or a given static spatial structure, Moran' I test is mostly adopted to detect the significance of

spatial autocorrelation.

$$\begin{aligned} I &= \frac{N}{W} \times \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \\ &= \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} = \frac{\sum_i (z_i \times \sum_j w_{ij} z_j)}{\sum_i z_i^2} \end{aligned} \tag{3.2.20}$$

where W is a row normalized spatial weight matrix, x_i is the variable of interest,$\bar{x}$ is the grand mean of x_i, therefore z means z-score as $z_i = \frac{x_i - \bar{x}}{SD(x_i)}$. In others words, Moran's I can be taken as a βslope estimator of regressing x_i on its spatial lag variable as $\sum_j w_{ij} z_j$.

2.panel data: As Moran' I is quite clumsy with the panel data, Lagrange Multiplier test was utilized to diagnose the CD's significance as bellow (Pesaran 2004):

$$\begin{aligned} CD_{lm} &= T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}^2 \\ \hat{\rho}_{ij} = \hat{\rho}_{ji} &= \frac{\sum_{i=1}^{T} e_{it} e_{jt}}{(\sum_{i=1}^{T} e_{it}^2)^{1/2} (\sum_{j=1}^{T} e_{jt}^2)^{1/2}} \end{aligned} \tag{3.2.21}$$

where e_{it} is the estimation of u_i by estimating OLS model that $e_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}_i' x_{it}$ for each i.And based on the regression of e_{it} on e_{jt} and maximum likelihood estimation, the statistic above can be derived. And under the assumption of cross-section independence,hypothesis H_0 can listed as $T\hat{\rho}_{ij}^2 \overset{a}{\sim} \chi_1^2$ with pairwise $\hat{\rho}_{ij}^2$ being asymptotically independent as $T \rightarrow \infty$.Therefore the static CD_{lm} in 3.2.21 is asymptotically distributed as chi-squared with $N(N-1)/2$ degrees of freedom if T is large enough.

Nevertheless statistic CD_{lm} in 3.2.21 is probably in distortion to handle the large N and short T panel data. In our panel data T is just 16, so in order to accommodate the short panel another more general statistic (Pesaran 2004) is adopted here.

$$CD_{lm}^* = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{t=1}^{T} \hat{\xi}_{it} \hat{\xi}_{jt}}{\{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{t=1}^{T} \hat{\xi}_{it}^2 \hat{\xi}_{jt}^2\}^{1/2}} \tag{3.2.22}$$

Different with 3.2.21, $\hat{\rho}_{ij}$ is calculated with the scaled residual $\xi_{it} = \frac{e_{it}}{e_{it}' e_{it}}$ as $\hat{\rho}_{ji} = \frac{\sum_{i=1}^{T} e_{it} e_{jt}}{(\sum_{i=1}^{T} e_{it}^2)^{1/2} (\sum_{j=1}^{T} e_{jt}^2)^{1/2}} = \sum_{i=1}^{T} \hat{\xi}_{it} \hat{\xi}_{jt}$. In this general form unbalanced panel and a priori spacial weight matrix can be accommodated here. Especially the row-normalized neighbour matrix can also be adopted here just as it is, when dependence is only diagnosed among the current site i and its neighbours j.

However in order to get the unbiased and consistent result, the following assumptions must be satisfied for the panel data model adopted:

(1)For each i the disturbance u_{it} are serially independent with zero means and the variance σ_i^2 that $0 < \sigma_i < \infty$.

(2)Under the null hypothesis that $H_0 : u_{it} = \sigma_i \epsilon_{it}$ with ϵ_{it} is in an independent and identical distribution $IID(0,1)$ for all i and t, the distribution of ϵ_{it} is symmetrically distributed around 0.

(3)The regressors x_{it} are strictly exogenous such that $E(u_{it}|X_i) = 0$ for all i and t where $Xi = (x_{i1}, x_{i2}, \ldots, x_{iT})'$ and $Xi' Xi$ is a positive definite matrix.

(4)Longitude $T > k + 1$ and the OLS residuals e_{it} are not all zero.

3.2.2 Regression models

Moineddin,et al.(2003) assumes that the probability to get engaged in a sector are independent across all the subregions. Obviously, this assumption doesn't hold for the scenario of location choice by firms, as firms used to consider competition, accessibility to upstream and downstream industries, traffic, scale, etc of the available location choices. Here I mainly focus on the way to estimate the correlation between the firm's scale with proxy variables about the factor mentioned in each subregion. In most literatures the dependant variables are only influenced by current subregion's regressors, which neglect the effect by other subregions' variables or residuals.

For observations about cities it's very attractive to detect whether and to what extent, the dependant variables are affected by other subregions.Therefore the regression models here must take the cross-section correlations into consideration.

Firstly FE model(i.e. fixed effect model), RE model(i.e. seemingly unrelated regression with pre-assumed error structure), fixed effect generalised least squares model(FEGLS) and feasible general least squares model (FGLS) are adapted here as the basic model. And it is still very important to stress their resumptions beforehand. These 4 models can be listed in the same structure as bellow:

$$y_i = X_i\beta + c_i e_T + u_i \tag{3.2.23}$$

,where y_i is a T×1 vector, X_i is a T×K matrix, e_T is a T×1 vector of ones, u_i is a idiosyncratic error and c_i are time-invariant variables which are taken as parameters in FE and FEGLS models, but exogenous stochastic variable in RE and FGLS models. In addition the assumptions is summed up as bellow:

1. FE model
(1)$E(\ddot{u}_{it}|x_i) = E(u_{it}|x_i) - E(\bar{u}_{it}|x_i) = 0$,
(2)$rank(\sum_{t=1}^{T} E(\ddot{x}_{it}^{'}\ddot{x}_{it}))$=$rank(E(\ddot{X}_i^{'}\ddot{X}_i)) = K$,
(3)$E(u_i u_i^{'}|x_i, c_i) = \sigma_u^2 I_t$;

2. FEGLS model
With the existence of the heteroskedasticity and serial correlation, the third pre-assumption of fixed effect model is relaxed as $E(u_i u_i^{'}|x_i, c_i) = \Lambda$, a $T \times T$ positive definite matrix, which can be Cholesky decomposed.Although the specific distribution of u_{it} and the serial correlation are agnostic, the estimator of β can still be consistent and unbiased. Accordingly the second assumption of FE model is updated as $rank(E(\ddot{X}_i^{'}\Omega^{-1}\ddot{X}_i)) = K$, where $\Omega = Q^T\Lambda Q$ and Q^T is a demeaning process to get rid of time-invariant individual effect c_i;

3. RE model

(1)$E(u_{it}|x_i, c_i) = 0$,

(2)[10]$E(c_i|x_i) = E(c_i) = 0$,

(3) rank$E(X_i^{'}\Omega^{-1}X_i)$=K,

(4)$E(u_i u_i^{'}|x_i, c_i) = \sigma_u^2 I_T$,

(5)$E(c_i^2|x_i) = \sigma_c^2$. So for each object i, its idiosyncratic errors $E(v_i) = E(c_i) + E(u_i)$ have a special homogeneous form as:

$$\Omega = E(v_i v_i^{'}) == \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \vdots & \ddots & \ddots & \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \dots & \dots & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

4. FGLS model

Just as the relaxation of idiosyncratic error's assumption,$\Omega = E(v_i v_i^{'})$ which can be estimated with the residuals of pooled OLS residuals. And only if the exogenousity assumption in RE effect is not violated, FGLS's estimator is consistent unbiased and more efficient than FEGLS model.

With the existence of cross-sector dependence, the basic regression models can be extended to spatial models in order to contain it. Manski (1993) points out that three different types of interaction effects can explain why an observation associated with a specific location can be dependent on observations at other locations: (i) endogenous interaction effects, where the decision of a spatial unit (or its economic decision makers) to behave in some way depends on the decision taken by other spatial units; (ii) exogenous interaction effects, where the decision of a spatial unit to behave in some way depends on independent explanatory variables of the decision taken by other spatial units - if the number of independent explanatory variables in a linear regression model is K, then the number of exogenous interaction effects is also K, provided that the intercept is considered as a separate variable; and (iii) correlated effects, where similar unobserved environmental characteristics result in similar behaviour. Therefore A general nesting spatial(i.e. Manski model) model for N sections with a time span of T can be listed as bellow:

$$\begin{aligned} Y &= \rho W Y + \alpha \iota_N + X\beta + W X \theta + \mu \\ \mu &= \lambda W \mu + \epsilon \end{aligned} \tag{3.2.24}$$

Where Y denotes an $N \times 1$ vector consisting of one observation on the dependent variable for every unit in the sample ($i = 1, \dots, N$), ι_N is an $N \times 1$ vector of ones associated with the constant term parameter α, W denotes an $N \times N$ spatial weights matrix, X denotes an $N \times K$ matrix of exogenous explanatory variables with the associated parameters β contained in a $K \times 1$ vector, the variable WY denotes the endogenous interaction effects among the dependent variables, WX the exogenous interaction effects among the independent vari-

[10]besides the same variables in FE effect model, a shared intercept in included in RE model

ables, and $W\mu$ the interaction effects among the disturbance terms of the different spatial units.ρ is called the spatial autoregressive coefficient, λ the spatial autocorrelation coefficient, while θ as for β, represents a $K \times 1$ vector of fixed but unknown parameters.

It seems very reasonable to adopt this general model, but J.Paul Elhorst(2010) proved that although the 3 types of interaction can be estimated technically the endogenous and exogenous effects can not be distinguished from each other. Therefore the main spatial panel models adopted here include two categories which impose restrictions on the interaction parameters ρ, θ respectively, alongside with existence of λ. The reason is that significant cross section dependence exists with basic model's regression results.

1. SAC(spatial autoregressive combined model)

$$\begin{aligned} y = X\beta + \alpha\iota_{NT} + \rho(I_T \bigotimes W)y + u = Z\gamma + u \\ Z = (X, \iota_N, (I_T \bigotimes W)y), \gamma^{'} = (\beta^{'}, \alpha, \rho) \\ u = ((I_T \bigotimes \lambda W)u + \epsilon \\ \epsilon = (e_T \bigotimes I_N)\mu + v \end{aligned} \tag{3.2.25}$$

2. SDEM(Spatial Durbin error model)

$$\begin{aligned} y = X\beta + (I_T \bigotimes W)X\theta + u = Z\gamma + u \\ Z = (X, (I_T \bigotimes W)X), \gamma^{'} = (\beta^{'}, \theta^{'}) \\ u = ((I_T \bigotimes \rho_2 W)u + \epsilon \\ \epsilon = (e_T \bigotimes I_N)\mu + v \end{aligned} \tag{3.2.26}$$

$E(\epsilon) = \mu$ and $var(\epsilon) = v^2$ in these 2 models with X as a $NT{\times}k$ matrix for T times' observation about k variables of interest, ι_{NT} is a $N{\times}T$ vector of ones.

Based a spatial lag two stage least squares estimator(weighted 2 stages least squares for the fixed effect here) by H.Kelejian（2004） on the these two models the spillover effect can be inferred by the direct effects and indirect effects. For SAC model,they can be derived as bellow(chapter2 Lasage, 2008):

$$\begin{aligned} (I_{NT} - \rho(I_T \bigotimes W))y = X\beta + \alpha\iota_{NT} + u \\ y_t = \sum_{r=1}^{k} S_r(W)x_{rt} + V(W)\iota_n\alpha + V(W)u_t \\ S_r = V(W)I_N\beta_r \\ V(W) = (I_N - \rho W)^{-1} = I_N + \rho W + \rho^2 W^2 + \rho^3 W^3 + ... \end{aligned} \tag{3.2.27}$$

,where y_t is $1{\times}N$ vector of tth time observation. x_{rt} is $1{\times}N$ vector of tth time's observation about rth variable of interest. And the measured impacts is defined as bellow:

$$\begin{aligned} \bar{M}(r)_{direct} = n^{-1}tr(S_r(W)) \\ \bar{M}(r)_{total} = n^{-1}\iota_N^{'}S_r(W)\iota_N \\ \bar{M}(r)_{indirect} = \bar{M}(r)_{total} - \bar{M}(r)_{direct} \end{aligned} \tag{3.2.28}$$

Moreover the direct effect $\bar{M}(r)_{direct}$ not only include the immediate direct effect (I_N in $V(W)$) but also the feedback effect by its neighbours.(diagonals of $\rho W + \rho^2 W^2 + \rho^3 W^3 + ...$). The off-diagonals' sum measures the indirect effects which numerically equals to emanating effects[11](Hondroyiannis et al. 2009) or vulnerability effects, which is the row's sum.[12] As to the SDEM model, direct impacts and indirect impacts are explicitly measured by β and θ respectively.

3.3 Data and variables

3.3.1 Data

Yearly localization quotient of function 3.2.5 will be applied to the ASIE database using the subtotal employee number. Without using the arbitrary threshold value of one,the confidential interval can be calculated based on the method of Moineddin(2003).

To be compared with the discrete agglomeration indices, employee weighted global localization and dispersion indices based on $K(d)$ function is calculated. And it can be inferred about the difference between the outcomes of these two categories' indices.Especially all the geographic information is scraped from the google map API[13] based on the zip codes attached to ASIE.

According to the zip code composing rules by Chinese post system, the first 2 digits represent the provincial administration division, the 3rd and 4th digits represent the prefectural-level and county-level administration division,and the last 2 digits represent the delivery office. Therefore it is achievable to scrap the delivery offices' geographic information. To calculate the mutual distance, all the scraped latitudes and longitudes are projected to euclidean space with coordinates reference system set as WGS84.

As to the specific zip codes' data process in ASIE, inaccurate entry is a serious matter that some zip codes are contaminated by abnormally excessive entries with last 2 digits as zeros. Indiscreetness with these raw data can cause serious biasness. To moderate this biasnees interpolation is utilized by replacing the erroneous data entries with time serially nearest zip code entries. Furthermore each firm can only hold a unique position in kernel distance function according to Duranton and Overman（2004）,therefore each firm's coordinate are estimated by jitterring the delivery offices' coordinates. In addition jitter deviates are less than the half of minimum distance among all zip codes' coordinates. This ad hoc method probably generates a systematic

[11]a change in a regressor relating to a given unit fan out to all the units, which equals the column's sum $S_r(W)$.

[12]Vulnerability describes the response of a given unit to the neighbouring units

[13]https://developers.google.com/maps

error but can simplify the computation process significantly.

To get free from any distribution presumption, I try to resort to other robust methods to estimate the discrete location measurement as well as the scale measurement. The most intuitive way is to adopt the kernel density estimation, but the ASIE data is exhaustive with regards to cities and the pre-decided scope's firms. So the distribution of LQ or SLQ can only be the forms of probability mass distribution. Moreover according to the function form of LQ values, each subregion's LQ is correlated through the denominator for the given industry. And it can cause systematic errors to bootstrap the LQ or SLQ values directly.Therefore I am restrained from estimating the robust statistics about the LQ values. In ASIE a 4 digits industry classification is provided for the level-1 categories of mining,manufacture and public services. We only focus on the manufacture sector as the other two are more vulnerable to the unevenness attributes of each city.As to the manufacture sector, we mainly focus on the 2 digit classification. In addition, NBS have amended the industry classification standards twice during year 1998-2013[14] ,which is divided into 3 stages.To unify the classification, I transfer all the industry codes uniformly to the GB/T 4754—2002[15] according to the cross-reference tables by NBS.[16] With respect to the scope of cities here, it refers to 332[17] prefectural-level units, 4 provincial-level cites and 8 other spatial units[18]

[14]GB/T 4754 — 1994 for 1998-2002, GB/T 4754—2002 for 2003-2012 and GB/T 4754—2011 for 2013

[15]Appendix Table 7.1

[16]http://tjj.beijing.gov.cn/zwgkai/tjbz_31390/xyhcyfl_31392/gmjjxyfl_31675/202002/t20200214_1631921.html

[17]Sansha(三沙) of Hainan province isn't included here

[18]This 8 spatial units include 2 groups: Laiwu(莱芜)，Xiantao（仙桃），Tianmen（天门）,Shennongjia(神农架)，Qianjiang(潜江)，Jiyuan(济源) and Shihezi(石河子), all of whom are county level cities directly under governance of province; Hainan province's left-out area by Haikou, Sanya and Danzhou.

Figure 3.1: Main cities in Chinese mainland

3.3.2 Variables

For any given industry A,linear models for N's subregions and T's periods can be listed as bellow:

$$q^s_{i,A} = F(q^x_{ir}, competition^A_{i,t}, diversification^A_{i,t}) \tag{3.3.1}$$

, where $q^s_{i,A}$ and q^x_{ir} are defined as function 3.2.6; $competition^A_{i,t}$ and $diversification^A_{i,t}$ are measurements about the competition and the diversification for current industry A in subregion i. And both of these are expended based on the Herfindal index. The index to measure competition is as bellow:

$$competition^A_{i,t} = ln(\frac{1}{HHI^A_{i,t}}) \tag{3.3.2}$$

$$HHI^A_{i,t} = \sum_{f \in S^{f,A}_{i,t}} (\frac{E_{f,t}}{E^A_{i,t}})^2$$

,where $S^f_{i,t}$ means all the firms that belong to the subregion i' firm group of current industry, $E_{f,t}$ means firm f's number of employees in period t, and $E^A_{i,t}$ means subregion i's number of A industry's employees in period t.

Likely another excluded competition index can be achieved to measure the competition level of current firm f:

$$competition_{i,t}^{f,A} = ln(\frac{1}{HHI_{i,t}^{f,A}}) \tag{3.3.3}$$

$$HHI_{i,t}^{A,f} = \sum_{j \in S_{i,t}^{\bar{f},A}} (\frac{E_{j,t}}{E_{i,t}^{A}})^2$$

The other index to measure industry diversification is as bellow:

$$diversification_{i,t}^{A} = ln(\frac{1}{H_{i,t}}) \tag{3.3.4}$$

$$HI_{i,t}^{A} = \sum_{\bar{a} \in \bar{A}} (\frac{E_{\bar{a},t}^{j}}{E_{i,t} - E_{i,t}^{A}})^2$$

,where $\bar{A}$ means all the manufacture sectors other than A that belong to the subregion i, $E_{\bar{a},t}^{j}$ means the count of $\bar{a}$ sector's employees in period t and subregion i, and $E_{i,t}^{A}$ means subregion i's count of A industry's employees in period t. To contain the spatial autocorrelation factor, based on the models mentioned above a spatial weight matrix is added in, which is constructed on the neighbouring relation among the cities in China. Through this, it is aimed to detect the spillover effect which includes emanating effect and vulnerability among the cities.Especially emanating effect means how much the current city's one unit change will affect all the other cites and vulnerability effect means how a city reacts to a uniform worsening(decrease) in all other cities.

Besides $competition_{i,t}^{A}$ and $diversification_{i,t}^{A}$, other control variables are introduced as bellow:
$City\_ind\_E$: each sector's yearly total number of employees in city i.For missing entries of data, interpolation is adopted here.
$City\_E$:yearly total of manufacture sectors' employees in city i.
$CusumYr\_pat$:one year lagged cumulative number of invent patents in city i, which is based on the data from CNKI.

3.4 Results

3.4.1 Result of the agglomeration

Firstly,with the function 3.2.1,3.2.3 and 3.2.4 LQ, SLQ and SLLQ values for each city across year 1998-2013 can be calculated. According to the derivation by Moineddin,et al(2003) there is a dichotomy about whether

$p_{i,A} = p_A$ across the whole nation. Nevertheless, the normal distribution test about $\frac{E_i \times p_{i,A} - E_i \times p_A}{E_i \times p_A \times (1-p_A)}$ is denied for all the industries. Therefore only function 3.2.2 for heterogeneity can be adopted for ASIE. Thereafter based on each city's confidence interval for all industries during the period span,those cities with confidence interval over or bellow one is taken as statistical significantly cites with concentration or dispersion for focal industry. The result can be summed up as figure 7.39 shows. In addition shapiro-wilk test is adopted to test the normality of LQ,SLQ and SLLQ for each sector yearly. And nether LQ, SLO nor SLLQ are in normal distribution. Therefore we cannot directly take the αor 1-αpercentile as the cut-off values. According to the results of the yearly shifts(Figure 7.39), thirty 2-digit manufacture sectors can be categorized into 4 groups: Sectors in fields of food and drink manufacturing(13, 14, 15), apparel & footwear(18), chemical raw material & pharmaceutical(26 ,27), non-metallic(31), ferrous metals' process(32), petroleum (25) and waste processing(43),more employees are concentrated in limited cities. In addition out of the 344 cities,the sum of concentration and dispersion cities outnumber the count of randomly distributed cites for nearly all the sectors except sector 16(tobacco). however it should still be emphasized that the presumption is far away from reality especially that the geographic unevenness and spatial correlation is not excluded here.

Table 3.1

Group	Characristics	Sector
Concentration Enhanced	Count of concentration cities mount up and count of dispersion cities level off or drop down	13,14,15,18, 26,27,31,32
Dispersion Enhanced	Count of dispersion cities mount up and count of concentration cities level off or drop down	17,20,21, 24,28,40
Homogeneity Enhanced	Both counts of dispersion and concentration drop down	16,23,37
Polorization Enhanced	Both counts of dispersion and concentration mount up	25,43
Others	Without significant change	19,22,29,30,33,34, 35,36,39,41,42

To test the spatial correlation hypothesis, the geographic layout of LQ values for year 1998 and 2013 are depicted as figure in section 7.2. From the figures' result, it seems that high LQ valued cities are more inclined to neighbour each other. To clarify the spatial autocorrelation, Moran' I test is be carried out for each year-sector separately, whereas the longitudinal correlation is not involved here as Moran's I test is not appropriate for the panel data. As to the spatial weights matrix W, a queen-type neighbour based on the Chinese administrative map is generated. If city i and j are neighbours, $W_{i,j}$ is valued as one. Otherwise, it is valued as zero.Secondly each row is normalized as $\frac{W_{i,j}}{\sum_j W_{i,j}}$,so sum of each row equals one. As to the cities that have no employee enrolled in the ASIE, their LQ values are all set as zeros.

In table 7.2 year 1998, 2006 and 2013's data are sliced to give a glimpse about the Moran's I two-sided test. Except the bold font figure all the p-values is extremely significant against the hypothesis that firms are randomly distributed. And among these sectors 18, 24, 40 are of with most significance.

Compared with the LQ's result, only sectors of furniture, ferrous medal processing and electrical machinery(21, 28 and 39) exhibit obviously strengthening trends of concentration, according to the global localization index in figure 7.37 and 7.38 ,which are based on the result of function 3.2.19[19]. Especially based on the longitudinal shift with respect to global kernel density result(section 7.4), sectors of food & agricultural product(13, 14), beverage(15), paper & printing(22, 23), chemical raw material & pharmaceutical(26, 27), non-metallic product, medal processing(31, 32, 33), and transport equipment(37[20] hold comparatively stable pattern that: within about 500 kilometres of diameter same sector firms are distributed scarcely, but there are multi clusters keeping over 1000 kilometres away from each other; sectors of tobacco(16) are of random distribution; in sectors related to apparel and footwear (17, 18, 19) firms keep a stable distribution with one cluster; in sector of timber wood product(20) multi clusters nationally get more adjacent resulting in one cluster; in sector petroleum(25) and rubber product(29) multi clusters emerged nationally; in plastic and artwork sectors (30, 42) one cluster evolved into multi clusters; in non-metallic mineral product sector (31) firms keep excessive distances away from their peers with the multi clusters distancing away further; in ferrous metal processing(32) dispersion within 500 kilometres dissolved but without significant agglomeration; in non-ferrous metal processing sector(33) multi national clusters shorten the distances among them, meanwhile firms get concentrated within each cluster; in sectors of education & art product, chemical fibers and metal products(24, 28, 34) firms get more concentrated into one cluster; in general purpose machinery sector(35) there was one stable cluster; in furniture & special purpose machinery sectors (21, 36) firms get more adjacent to their peers within 500 kilometres with multi clusters emerging with distances of range 800-1000 kilometres; firms of electrical machinery and communication equipment(39, 40) were densely distributed within 500 kilometres with multi clusters keeping 1000-1500 kilometres away form each other; firms of measuring instruments sector (41) had a pattern like sector 39 and 40 except late stage during which the multi clusters were integrated as one. At last in sector of waste processing(43), the pattern is very unstable partly because NBS just enlisted this sector from 2002.

3.4.2 Result of the basic regressions

Based on the Hausman tests in table 7.32 and 7.33, RE and FGLS models are not consistent. Meanwhile there are also strong serial correlation and cross section correlation with respect to the residuals of the regression for each sector. Therefore FE and FEGLS models' outcomes are consist compared with the other 3 models,

[19]60 simulations are carried out for each sector yearly with the package of dbmss in R language.

[20]different from others firms of transport equipment got more concentrated in each cluster since 2011

and FEGLS is robust to intra-group heterogeneity and serial correlation.

To sum the results, the scale of focal sector q_{ir}^x(i.e. $LQ\_general$) are positively correlated with comparative average firm size high q_{ir}^s significantly in all sectors and the degree of competition are all significantly negative related with the q_{ir}^s. Meanwhile diversity of manufacture sector have significantly negative effect except sector 16(tobacco product) and 43 (recycling and disposal of wastes). As for the effect of focal sector's absolute scale($City\_ind\_E$), it is found that swelling sector's size will uniformly contribute to increasing firm's size. Therefore large scale firms of all sectors significantly prefer to be localized in the sites with high Marshallian externalities based on the positive correlation with $LO\_general$ and $City\_ind\_E$. On the contrary small sized firms are more attracted by Jacobian across sector diversification externalities except tobacco producers and waste processors. Moreover for the traditional sectors like tobacco, apparel, leather, timber, plastic , metal products and waste processing, large sized producers have the significant propensity to distance away from cities with large scaled composite manufacture sector($City\_E$). The only exception is with electrical machinery and equipment sector (39) that large sized firms also tend to distance form large cities. As to the competitiveness, it is compliant with the traditional economic theory that with the existence of high competitiveness firms tend to have little influence on the market, which is manifested by the comparative small size of firms.

In addition, the effect of city's cumulative patent count have diverging effects that firm's size is affected positively in sector 17, 20, 22, 25, 26, 28, 29, 31, 33, 35, 36 and 42. Coreference with the industry classification, in the sector of textile, timber, paper, petroleum, chemical raw material, chemical fiber, rubber, non-metallic mineral product, non ferrous metals, machinery and artworks, other than generating more firms the existing firms tend to grow bigger. For the other sectors the growing patents contribute to more small sized firms participating into the focal sector.

3.4.3 Result of the spatial econometric regressions

According to tables in section 7.2, the cross-section dependence is overwhelming across all sectors. To better explain the spatial spillover mechanism among the cities, a KNN(kernel nearest neighbour) spatial matrix is introduced taking the longest distance between each city' geometry centre with its nearest neighbour's as the threshold value.Thereafter all the cities with geometry centre within the threshold value's distance for focal city are taken as its neighbours. And the the weight is taken as inverse value of the distance between it and its neighbours. Thereafter this spatial weight are row standardized. Other than queen type the reasons to adopt KNN-matrix is that not all cites hold all sectors and there are many isolated cities whose bordering cities hold no firms for focal sector.

Table 7.37, 7.38, 7.42 and 7.43 in section 7.3 sum the basic results from SAC models and spatial Durbin error models.[21] Especially based on the regression results of SAC model, the direct and indirect impact result are derived in Table 7.40 and 7.41 according to function 3.2.28 where the significance is inferred based on 200 simulations with the fitted models.[22]. Different from the basic model spatial econometrics models explicitly separate direct effect from the indirect effect for the variable of interest.

With both of the spatial econometric models consistent with basic models, localization quotients contributes to increasing size of firms except sector 24(education, art and sports products). Comparing the results of the two spatial econometric models, the positive indirect effects of localization exist dominantly in sectors of machinery(35 & 36), metal products(34), non-ferrous metal processing(sector 33), plastic(30), non-metallic mineral products(31), petroleum(25), beverage & alcohol (15) and tobacco(16).For the remained sectors it is not deterministic with the indirect effects of localization.

Meanwhile focal sector's absolute size have deterministic positive indirect effect on neighbouring cities' firm size in petroleum (40), chemical materials(26), non-metallic mineral products(31), metal materials(34), special purpose machinery(36), tobacco(16), measuring instrument(41), artwork(42) and waste processing(43).

Consistently with the basic models Jacobian across sector diversification externalities also have negative direct affects on focal sector firms' size dominantly except sectors of tobacco, leather & fur and waste processing(16, 19, 43). As to the indirect effect, negative indirect effects can be ascertained in sectors of petroleum (25), chemical materials(sector 26), non-metallic mineral(sector 31), metals smelting & processing(sector 31,32) and general purpose machinery(35). With respect to the intra sector diversity or competition, the negative direct effect is consistent with the basic models. Although the indirect effect can be found in both models, the only deterministically negative effect exists in transport equipment sector and the deterministically positive effect exist in measuring instruments and artwork sectors(sector 41 and 42).

Different with the basic models, no negative direct effect of composite manufacture sector can be ascertained. Meanwhile positive direct effect can be found in most sectors except the traditional sectors mentioned for basic models alongside with sectors of electrical machinery and equipment sector(39), food and processing of agricultural products.

As to effects of patents' stock, based on both models it have positive direct effects on petroleum sector's firm size and negative direct effects on sectors of plastic product, smelting and processing ferrous metals, transport equipment and measuring equipment. With respect to the indirect effect, it has ascertained positive positive indirect effects in sectors of textile mills, chemical raw materials, non-metallic mineral products and measuring equipments.

The mentioned findings are all based on the double confirmed results by both SAC and SDEM models. For the other results, they should be compared based on the goodness of fit.Nevertheless there are still no decisive non-nested tests for the SAC panel and SDEM models.

[21]variables with "SL" are spatial lag of focal variables

[22]The calculation is done by spdep package of R.

Chapter 4

Influence of R & D network on firms' activity

4.1 Introduction

Cohen and Levinthal (1990) argue the ability of a firm to recognize the value of new, external information, assimilate it and apply it to commercial ends is critical to its innovation capabilities. Although they concluded that firms are sensitive to the learning environment and absorptive capability is in firm's decision calculus, it is intangible .

Gautam Ahuja(2000) elaborates a theoretical framework-direct ties, indirect ties and structural holes to assess the effect of a firm's network on innovations, which generate a frameworks to probe into absorptive capability. Especially they find increasing structural holes is associated with the reduced innovation output(i.e. granted patents).

Sherzod Aktamov and Yan Zhao(2014) attempt to find the relationship between network centrality indices and innovation performance with the evidence form Chinese auto-mobile industry. The network they constructed are based on the strategic alliance relation centering on the targeted auto-mobile firms. And they found well connection (eigenvector centrality) and direct influence(degree centrality) have a significant and positive effect on innovation(i.e. number of granted patents).

Jingjing Zeng, et al（2019）examines the agglomeration effects of industry cluster on firm's innovation performance through the network embeddedness of pharmaceutical companies in Wuhan. The network constructed is measured by inter-enterprise contracts or agreements. And they concluded that betweenness centrality and clustering coefficient have statistically significant and positive effects on enterprise's ability for technological innovation, while the influence from the constraint of structural holes is negative (i.e. maintaining more structural holes is preferable).

With respect to the agglomeration's effect on firms' innovation performance(launching new product), Zhang(2015) finds that diversity externalities are preferred to localization externalities. To complement the explanation,total factor productivity by Levinsohn and Petrin(2003) is firstly calculated for the enterprises in ASIE.

The existing research about Chinese innovation network mainly focus on the industry-specific regional networks. And the proxy data for the collaboration relation used is limited in comparative closed environment, which is very difficult to be ascertained for foreign researchers. Different from the existing research, it is targeted to probe into the effect of the overall co-patenting network on the firm-level performance alongside with the agglomeration. Co-patenting network can be taken as a token of the R& D network, tangibilizing mutual relationships among organization. Especially besides the new product output and wage-benefit, it is initiated to diagnose the firm-level estimated production efficiency's relation with co-patenting network, citation network and agglomeration.

4.2 Literature

A network is a set of objects (called nodes or vertices)that are connected integratedly. The connections between the nodes are called edges or links. With respect to direction of edge network can be categorized into directed network and undirected network. Without considering the number of edges between two nodes the network is labelled as unweighed network otherwise weighed network.For example figure 4.1 is a weighted undirected network and figure 4.2 is a unweighed directed network. In figure 4.1, the thickness of the edges represent the weights (i.e number) of edges between the connected two nodes. In figure 4.2, the edge stem from one node to another one, but not vice versa necessarily.

4.2.1 Network related statistics

Statistics related to co-patenting network

Degree of node v is the number of edges incident on it in the undirected network.For directed graphs the indegree of a node is the number of edges leading into that node and its outdegree is the number of edges

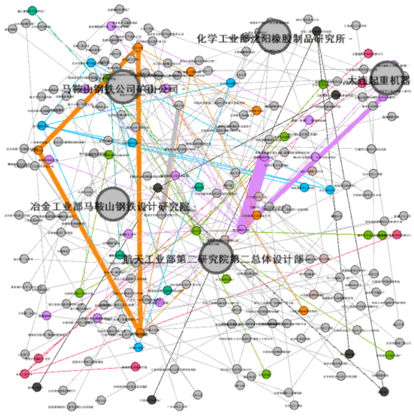


Figure 4.1: Co-patenting Network centering on Chinese firms by 1985

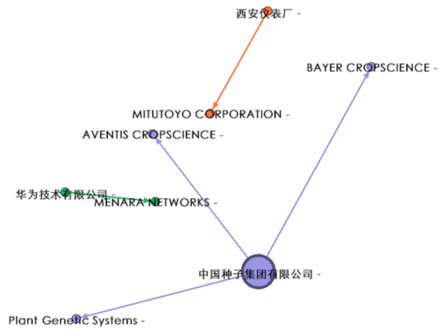


Figure 4.2: Citation Network centering on Chinese firms by 1990

leading away from it to its neighbours(i.e. alters)

Betweenness Centrality of node v: $C_b(v)$ is defined as bellow:

$$C_b(v) = \sum_{s \neq t \neq v} \frac{\theta(s,t|v)}{\theta(s,t)} \tag{4.2.1}$$

,where $\theta(s,t|v)$ means the number of paths between s and t that pass through node v among the shortest paths (i.e. geodesics) between s and t, whose count is labelled as $\theta(s,t)$. Betweenness centrality is a indicator of a node's capacity to connect different network subregions.

Burt’s constraint measure $Co(v)$ is defined as bellow:

$$C_o(v) = \sum_{s \in V_v, s \neq v} (p_{vs} + \sum_{t \in V_v, t \neq v, s} p_{vt} p_{ts})^2 \tag{4.2.2}$$

V_v is the set of nodes other than v in v's ego network, which means that V_v are all directly connected to the ego v. Consequently $C_o(v)$ is low when v is bridging various types of groups in the network. And $p_{vs} = 1/N_v$ is the relative link strength between nodes v and s,N_v is the degree centrality or number of link incidents upon node v. Moreover in weighted network $p_{vs} = N_{v,s}/N_v^w$, where $N_{v,s}$ is total number of link incidents between v and s and N_v^w is the number of total link incidents upon node v. So Burt's constraint index is used to measure v's connecting capacity in the neighbourhood region.

The higher Burt's constraint is, the more abundant (i.e. mutually stronger related) the contact is. Burt's constraint is calculated for each ego network of v respectively. To substantiate this algorithm, a simple example is as the following graphs shows.

The numeric attached to the edge means the weights. As both of the graphs are ego network, it means there

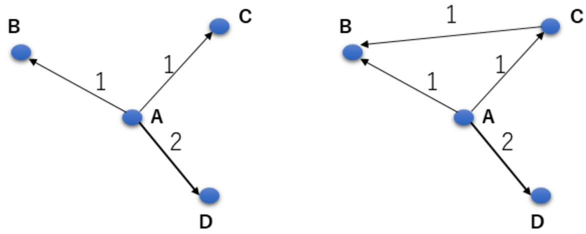


Figure 4.3: Ego Network of nodes A, B, C and D

are no other nodes connected to any of these four nodes. In the left network, B, C and D cannot share information except through node A. Nevertheless in the right one, B and C can share information directly without through node A who eclipses as the information bridge. Especially in the left one node A monopolize the information among them, in the right one the monopoly is break down between A, B and C. Burt's constraint is quite sensitive to probe into this difference. The node A's constraint on left equals $0.25^2 + 0.25^2 + 0.5^2 = 0.375$, compared with $(0.25 + 0.25 \times 0.5)^2 + (0.25 + 0.25 \times 0.5)^2 + 0.5^2 = 0.53125$ on the right one.

Citation network related statistics

Besides the statistics in 4.2.1, Hyperlink-Induced Topic Search(HITS) algorithms are adopted, which is originally developed by Jon M. Kleinberg(1999)to rate the web page's information during the forming process of internet. Especially to analyse the information source, each page can be estimated by the number of linking to other pages (hub) and the number of being linked by other pages (authority). This algorithm can also be applied to social network analysis with pages taken as nodes. In detail to calculate the hub and authority weights, firstly each node are assigned initial values as $1/\sqrt{N}$[23] for both the hub weight h_i^0 and authority weight a_i^0. From here on, all nodes $\{i\}$ embraced in the network will be given a updated authority weight $\{a_i^t\}$ and a updated hub weight $\{h_i^t\}$, which equal the sum of h_i^{t-1} of all nodes that points to i, and the sum of $\{a_i^{t-1}\}$ of all nodes that is pointed to by i respectively, where the superscript t mean the tth iteration. And this iteration will be carried out that both values converge.

To substantiate this algorithm let square matrix M be the adjacency matrix of a network with N nodes, where all diagonal entries of M are all zeros, element a_{ij} with entry of one or zero means whether there are links from node i to node j or not, and $a_{ij} \neq a_{ji}$ for all as links are directed. In addition vector v and u are authority weight vector and hub weight vector for nodes set $\{i\}$.

[23] All the weigh values are normalized that $\sum_i (a_i^0)^2 = \sum_i (h_i^0)^2 = 1$

$$A = \begin{bmatrix} 0 & a_{12} & \dots & a_{1N} \\ \vdots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots \\ a_{N1} & \dots & \dots & 0 \end{bmatrix}, v = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, u = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix}$$

The initial hub and authority weights for $\{i\}$ is assigned values as $u_0^t = v_0^t = [1/\sqrt{N}, 1/\sqrt{N}, ...1/\sqrt{N}]$ and iteration can be summed as:

$$\begin{aligned} \forall i : Authority : a_i^{t+1} &= \sum_{j \to i} h_j^t \\ \forall i : Hub : h_i^{t+1} &= \sum_{i \to j} a_j^t \\ \forall i : Normalize \sum_i (a_i^{t+1})^2 = 1, &\sum_i (h_j^{t+1})^2 = 1 \end{aligned} \tag{4.2.3}$$

Therefore it can be inferred that:

$$\begin{aligned} u = A \cdot v \qquad & v = A^T \cdot u \\ u^t = (A \cdot A^T) \cdot u^{t-1} \qquad & v^t = (A^t \cdot A) \cdot v^{t-1} \end{aligned} \tag{4.2.4}$$

Therefore the converged values of v and u are the eigenvectors of AA^T and $A^T \cdot A$ respectively.

Visualization and descriptive statistics about the networks

The Gelphi is utilized here to visualize the yearly edge lists parsed out, some years of which are shown by section 8.2 and 8.3. The overall descriptive statistics can be summed as bellow[24]. Both of the networks extended overwhelmingly from 1998 to 2013, especially after year 2008.Meanwhile the two networks became more sparse that density of both networks decline significantly.[25]

And top 10 organizations ranked with betweenness centrality and HITS yearly for co-patenting and citation networks are in table 8.1, 8.3 and 8.4. It can be found that the universities hold the central position in the co-patenting network alongside with some mega firms in oil, steel, electricity supply and electronic equipment sector. In contrast with co-patenting network, firms hold the dominant positions in citation work. Especially Chinese firms mostly absorb unilaterally for foreign firms, even though the trend slightly reversed in the late stage of the period. In addition there is a shift of industrial composition in the citation network that firms in semiconductor and electronic equipment overpass the oil and steel firms in the late stage. And also the

[24]Avg. Deg means average degree, Avg. W Deg means average weighted degree

[25]Co-patenting network's density declined from 0.001 to less than 0.001; Citation network's density declined from 0.004 to less than 0.001. The density is calculated based on the ratio of edges' count to theoretically maximum edges' count. So for a undirected network with N nodes the maximum is N(N-1)/2.

interaction between Chinese firms and Taiwan's firms shift to more bilateral that some Taiwan's firms turn to absorb the patents by their Chinese peers.

Table 4.1: Descriptive statistics about the co-patenting network

Year	Nodes	Edges	Avg. Deg	Avg. W Deg	Year	Nodes	Edges	Avg. Deg	Avg. W Deg
1998	2446	2018	1.650	2.576	2006	7437	7203	0.969	3.438
1999	2648	2216	1.674	2.974	2007	9058	9119	1.007	4.373
2000	2944	2524	1.715	3.430	2008	11397	11735	1.030	4.628
2001	3321	2875	1.731	3.715	2009	14555	15510	1.066	4.804
2002	3846	3352	1.743	4.209	2010	18033	20076	2.227	10.318
2003	4612	4094	1.775	4.585	2011	21942	25398	2.315	11.201
2004	5343	4823	1.805	5.090	2012	26628	32274	2.424	12.561
2005	6272	5962	1.901	6.078	2013	31732	41120	2.592	14.995

Table 4.2: Descriptive statistics about the citation network

Year	Nodes	Edges	Avg. Deg	Avg. W Deg	Year	Nodes	Edges	Avg. Deg	Avg. W Deg
1998	249	266	1.068	1.474	2006	10631	18658	1.755	4.312
1999	845	1055	1.249	1.914	2007	13618	25841	1.898	4.922
2000	1337	1704	1.274	2.073	2008	17252	34345	1.991	5.318
2001	2138	2804	1.312	2.324	2009	21696	45269	2.087	5.755
2002	3022	4197	1.389	2.523	2010	27178	60079	2.211	6.110
2003	4252	6251	1.470	2.818	2011	33063	78886	2.386	6.638
2004	5701	8936	1.567	3.251	2012	40255	102932	2.557	7.362
2005	8069	13378	1.656	3.858	2013	48088	132336	2.752	8.276

4.2.2 Hypothesis about the co-patenting network.

R&D does not only generate new information but also enhances the firm's ability to assimilate and exploit existing information(Wesley.Cohen, et al 1989). The new information may refers to technology, technique, managerial structure, etc. The information can be kept in form of patent, trade secret, experience and established practices. The ability to generate,assimilate and learn is intangible too. In the research frame here, the patenting and citation networks is adopted as the proxy for these unmeasurable variables here.

The economic meaning of organization's position in the inter-organizational network have drawn many researchers' attention since 1990s(Powell et al 1996). However it is not conclusive that what the specific effects of the role played by different network structure is on firms' performance theoretically. Researchers used to adopt empirical ways to test the hypothesis about the networks.(e.g. Gautam Ahuja 2000,Akbar Zaheer 2005)

The centrality of the dominant is measured by using Freeman(1997)' s concept of "betweenness". The important idea here is that an a node is central if it lies between other actors on their geodesics implying that to have a large betweenness centrality, the node must be between many other nodes via their geodesics. So the higher the betweenness-centrality is, the more probable that the focal node get in touch with the existing knowledge resource that facilitate its innovation ability. Meanwhile due to the advantageous position in the network, the more central firm's economic performance may gain advantage as well.

Hypothesis 1: Betweenness centrality of the firm is positively related to its production efficiency level.
Hypothesis 2: Betweenness centrality of the firm is positively related to its wage level.
Hypothesis 3: Betweenness centrality of the firm is positively related to the innovation performance.

Closed innovation system and open innovation system are two main innovation systems. In the past, many companies believed that as long as they invested more heavily in R &D than their competitors and protected their intellectual property from spilling over, they could innovate faster and more radically than competitors and hence sustain their competitive advantage. This paradigm of innovation is called closed innovation which requires the aggressive control of internal knowledge from leaking outside (Herzog &Leker, 2010). Nevertheless with more openness of existing information, open innovation system seems gain advantage over closed ones these days. Especially for those nodes bridging structure holes, they input more resources to build up the capability to assimilate knowledge from outsiders of the existing collaborators, which can bring in more outstanding innovation and economical performance. On the contrary nodes with higher constraint are more concentrated to strengthen the existing collaboration relations.

Hypothesis 4: Structural holes in the constraint are negatively related to a firm's production efficiency level.
Hypothesis 5: Structural holes in the constraint are negatively related to a firm's wage level.
Hypothesis 6: Structural holes in the constraint are negatively related to a firm's innovation performance.
As to the innovation performance, new product intensity and the current year's count of invent patent application are adopted to quantify it.

Different from the statistics embedded in the co-patenting network, authority and hub scores can act as the proxy of innovativeness and absorptiveness bibliometrically. In the classical economic scenario, firms with more advanced technology usually can reap higher productivity gains. Meanwhile their peers endeavor to narrow the technological gap. Therefore it is meaningful to test which type of R& D capability more crucial for the improvement of productivity.
Hypothesis 7: Authority scores are positively related to the production efficiency.
Hypothesis 8: Hub scores are positively related to the production efficiency.

4.3 Methodology

4.3.1 Estimate of efficiency

The main incentive for firms to improve the production efficiency or initiate the new product is to gain supernormal profit over their competing counterparts. To estimate the firm level efficiency the stochastic frontier model containing the firm level heterogeneity (G.Battese & J.Coelli 1995) is adopted here to estimate the efficiency for firm i in time t here as bellow:

$$\begin{aligned} Y_{it} = x_{it}\beta + E_{it} = x_{it}\beta + V_{it} - U_{it} \\ U_{it} = z_{it}\delta + W_u \\ t \in \{1, 2, \dots T\} \end{aligned} \tag{4.3.1}$$

where Y_{it} is the logarithm of gross output, x_{it} are logarithms the input of production along with the tth period $.\beta$ is the parameters needed to be estimated, V_{it} is assumed to be iid $N(0, \sigma_v^2)$ and U_{it} is non-negative random variables,associated with the technical inefficiency of production, W_u is defined by the truncation of normal distribution with zero mean and a homogeneous variance $\sigma_U{}^2$ so U_{it} is obtained by truncation (at zero) of the normal distribution $N(z_{it}\delta, \sigma_U{}^2)$ of firms over time. In order to contain heterogeneity of inefficiency distribution, z_{it} is taken as a vector of explanatory variables associated with the technical progress of firm i's production in t. And δ are the parameters to be estimated. Furthermore the Farrell efficiency $E(-U_{it})$ labelled as $BC95$ (G.Battese &J.Coelli 1995) can be derived based on the regression result with the inferred conditional density function of $f_{U|e_{it}}$.

$$\begin{aligned} f_{U|E=e}(u) = \frac{exp\frac{-1}{2}[\frac{(u-\mu_\star{}^2])}{\sigma_\star^2}]}{\sqrt{2\pi}\sigma_\star\Phi(\frac{\mu_\star}{\sigma_\star})} \\ \mu_\star = \frac{\sigma_V{}^2 z\delta - \sigma_U{}^2 e}{\sigma_V{}^2 + \sigma_U{}^2} \quad \text{and} \quad \sigma_\star{}^2 = \frac{\sigma_U^2 \sigma_V{}^2}{\sigma_U{}^2 + \sigma_V{}^2} \\ E(-z_{it}\delta - W_{it}) = E(-U_{it}) = \int_0^{+\infty} -u f_{U|E=e}(u)\, du. \end{aligned} \tag{4.3.2}$$

And Φ is the cumulative distribution of standard normal distribution. And time-effect τ is included both in the main part and the residual part which can reflect the Hicksian neutral progress and the logarithm of Farrell efficiency's linear change with time(G.Battese &J.Coelli 1995). Moreover according to Olsen, et al.(2011) the marginal effect of $z_{kit} \in z_{it}$ on the Farrell efficiency can be estimated as bellow:

$$\frac{\partial E[exp(-U)]}{\partial z_{kit}} = \frac{\delta_k(1-\gamma)exp(-\mu_\star + 1/2\sigma_\star^2)}{\Phi(\frac{\mu_\star}{\sigma_\star})} \times \\ (\frac{\phi(-\sigma_\star + \mu_\star \backslash \sigma_\star)}{\sigma_\star} - \frac{\Phi(-\sigma_\star + \mu_\star \backslash \sigma_\star)\phi(\mu_\star \backslash \sigma_\star)}{\sigma_\star \Phi(\mu_\star \backslash \sigma_\star)} - \Phi(\frac{\mu_\star}{\sigma_\star})) \tag{4.3.3}$$

Therefore the derived firm-level $BC95$ can be regressed on the patent network related variables, agglomeration related variables and other control variables, which can give us a glimpse into how R& D network and agglomeration affect the efficiency.

Moreover the main part form of model adopts the translog production as bellow:

$$\begin{aligned} \log(OutPut_{it}) = \beta_0 + \beta_1 \log(FxAsset_{it}) + \beta_2 \log(Employee_{it}) + 0.5\beta_3 \log^2(FxAsset_{it}) \\ + 0.5\beta_4 \log^2(Employee_{it}) + \beta_5 \log(Employee_{it}) \log FxAsset_{it}) + \tau \\ \text{and} \quad U_{it} = \log(CumuPat_{it}) + \tau + W_{it} \end{aligned} \tag{4.3.4}$$

In addition translog can be taken as a Taylor series approximation of CES (i.e. constant elasticity of substitution) function $Y = (\omega_1 X_1^\rho + \omega_2 X_2^\rho)^{1/\rho}$ as bellow:

$$\begin{aligned} & x = X_1/X_2 \\ & \log g(x) = \log(Y/X_2) = \rho^{-1}\log[\omega_1 e^{\rho \log(x)} + \omega_2] \\ & \frac{\partial \log g}{\partial \log x} = \omega_1 x^\rho/(\omega_2 + \omega_1 x^\rho) = a \\ & \frac{\partial^2 \log g}{\partial \log x^2} = \rho\omega_2\omega_1 x^\rho/(\omega_2 + \omega_1 x^\rho)^2 = 2\rho ab \\ & 2b = \omega_2/(\omega_2 + \omega_1 x^\rho) \\ & \quad \text{for} \quad x \xrightarrow{P} x_0 \\ & \log(Y) = \log(X_2) + \log g(x) \\ & \approx \log(X_2) + \log(x_0) + a\log(x) - a\log(x_0) + b[\log(x) - \log(x_0)]^2 \\ & \approx \alpha_1 log(X_1) + \alpha_2 log(X_2) + \alpha_3 log^2(X_1) + \alpha_4 log^2(X_2) + \alpha_5 log(X_1) log(X_2) \end{aligned} \tag{4.3.5}$$

The translog production takes the yearly net fixed assets and employee as the main input. Although intermediate product can be found in ASIE ,it is excluded because it causes the misspecified model that the estimated residuals is in right-skewed distribution (i.e. the U_{it} are all negative so all the firms are in full-efficient operation.) And logarithm of cumulative count of patent applications and τth year are taken to be z_{it}. All the calculation is carried out with the frontier package in R.

4.3.2 Poisson fixed effect model

To estimate the co-patenting network's effect on the count of patent application, Poisson fixed effect is adopted here(Baltagi, 2013). The model for firm i in year t takes the form as bellow:

$$\begin{aligned} Pr(Y_{it} = y_{it}|x_{it}) &= \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{y_{it}!} \\ \log \lambda_{it} &= \mu_i + x_{it}\beta \end{aligned} \tag{4.3.6}$$

where μ_i refers to the unobservable individual specific effects, just as the fixed effect panel model. Thereafter $E(y_{it}|x_{it}) = var(y_{it}|x_{it}) = \lambda_{it}$ and for continuous variable x_{itk} in x_{it} the marginal effect is $\frac{\partial E(y_{it}|x_{it})}{\partial x_{itk}} = \lambda_{it}\beta_k$. Although many researcher criticize this model for its pre-requisite $E(y_{it}|x_{it}) = var(y_{it}|x_{it})$, it is adopted as other models cannot bring about converging maximum likelihood estimation result for the data.

Besides Burt's constraint and betweenness centrality, one year lag and current year's R & D expenditure is included here, because ASIE only provides it in year 2005, 2006, 2007 and 2010. As R & D expenditure in 2008 and 2009 is missing, data of year 2010 is dropped.

4.3.3 Variables

All the variables involved are as bellow:
Dependent variables:
$BC95$: Farrell efficiency derived from stochastic frontier model (G.Battese &J.Coelli 1995)
$WageBenefit_{per}$:annual average wage plus benefits per capita.
$Newproductintensity$: The new product output ratio to the total output which is only available in year 1998-2007 and 2010.
$patentcount$: The current year's count of newly filed in patent application.

Control variables:
$Output$: priced adjusted total output of firm i in year t.
$Employee$:Total count of Employee of firm i in year t.
$CumuPat$:one year lag cumulative number of invent patents beheld by the firm i in year t.
$Ownership$: factor variables includes private(baseline), state-owned, collective-owned, HMT-owned and foreign-owned. I use the percentage of these 5 types' percentage of entity recorded in ASIE. Besides this way registration types are also recorded in ASIE, but they are more comparatively static than the share of entity(Nie Huihua, 2012).Specially private-owned is taken as baseline altogether with 1 as state-owned, 2 as

HMT owned, 3 as foreign-owned and 4 as collective-owned.

Independent variables:
$Burt'sconstraint$: one year lagged constrain value calculated as the function $4.2.2$.
$betcent$: one year lagged betweenness centrality calculated as the function $4.2.1$
$authority$ one year lagged authority as the function $4.2.4$
hub one year lagged hub as the function $4.2.4$
$localization$:$\log(E_t^{sec,city} - E_{i,t}^{sec,city} + 1)$ where sec is the 2 digits sector of firm i and city is where the firm i is located. Localization can be seen as a proxy of Marshallian externalities.
$urbdiver$:urbanization diversity of manufacture sector apart from mining sector, which is based on the Herfindahl – Hirschman Index. And higher the $urbdiver$ is the more diversified the manufacture sector of the focal city is in year t for the sector sec. $urbdiver$ can be taken as the Jacobian diversification externalities.

$$urbdiver_t^{sec,city} = ln(\frac{1}{H_i^{sec,city}}), H_i^{sec,city} = \sum\nolimits_{set' \neq sec}(\frac{E^{sec' city}}{E_t^{city} - E_t^{sec,city}}) \tag{4.3.7}$$

sec is the focal sector and sec' is the all other sectors apart from focal sector.
$urbanizationsize$ is the yearly total manufacturing sector's employees' count other than the focal firm i in focal city, which is calculated as $ln({E_t}^{city} - {E_t}^i + 1)$.
$competition$: yearly competition level of sector which firm i belonged to in the city where it is located. And the calculation is same as function $3.3.2$.
$R\&Dexp$: current year's R & D expenditure.
$R\&Dexp_{lag}$: last year's R & D expenditure.

4.4 Results

4.4.1 Overall results of the Farrell efficiency

The efficiency' results are derived from the regression of each sector. Due to the model's misspecification, the cumulative counts of patents is omitted for sector 32 and the year's variable is omitted for sector 33. In addition the main part of the model is misspecified for sector 43 which is recycling and disposal of waste. It can be reasonably inferred that the production process of sector 43 is at odds with other manufacturing sectors. In section 8.4 it summarizes the yearly distribution of efficiency of each sector, where we can find obvious trends of convergence in sector 13,14,15,16,18,19,20,21,23, 26,27,30,31,34,35,36,37,39 and 41. On the contrary trends of divergence can only be found in sectors of petroleum, non ferrous metals processing and communication equipment(25,33 and 40). As to the most actively patenting sector 40 (communication equipment, computers and other electronics equipment), the distribution is bimodal. Especially the firms like Huawei, ZTE boast the outstanding efficiency apart from their domestic peers.

The shift trends can be partly explained by the patents filed as the section 8.5 shows. Especially the marginal contribution effects by patent monotonically decrease along the time in most sectors except those of tobacco, petroleum, chemical fiber and non ferrous metals(16, 25, 28 and 33). Although the patents application outburst after 2008, the contribution to the production's efficiency don't show the trend proportionately except the sector 25 (petroleum, coking and nuclear fuel process). As to the most actively patenting sector 40 (communication equipment, computers and other electronics equipment), the distribution is bimodal. Especially the firms like Huawei, ZTE boast the outstanding efficiency apart from their domestic peers.

4.4.2 Results of the regression

The observations included in the regression are those enlisted in networks based on their stock of patent. Therefore the data is a monotonically incremental unbalanced panel data. The regression can only probe into the mechanism of the existing firms in networks, but not estimate under what conditions a firm has the will to stretch out for a co-patenting partner or assimilate others' patent.

Pooling, fixed effect(FE) and random effect(RE) model are adopted as basic models. According to the Hausman test between FE and RE the regressors are not exogenous to the estimated errors. Moreover the FE model is not compliant with the homoscedasticity and serial uncorrelated hypothesis, which means that the estimator is lack of efficiency. Therefore robust estimators are adopted including intra-group heteroscedasticity and serial correlation robust estimator(Arellano 1987) and FEGLS/FDGLS(First difference general least squares). Nevertheless the existence of cross section dependence dampens the efficiency of these two alternatives to the basic models because cross section dependence implicates that covariance matrix of the residuals are not identical among the the groups(firms). Despite the heterogeneity of each firm, it is still reasonable to take the results of FE and FEGLS models as consistent.

As sections 8.6, 8.8, 8.9 and 8.10 show, Burt's constraint is negatively related with the technical efficiency, wage benefit level, new product intensity and patent count with strong significance. On the other side betweenness centrality has only significantly positive affect on patent count, meanwhile its positive effect on Farrell efficiency is questionable. It can be inferred that betweenness in the total graph can not bring about direct effect on firms' business performance, but it can subtly affect it by boosting firm's patent application and stock of knowledge. On the other side Burt's constraint not only boost the patent application but also affect firms' business performance directly.

With respect to the effect of the agglomeration externalities, scale related variables(localization and urbanization size) outshines the inter-sector diversity externalities for the Farrell efficiency. Even urbanization diversity show to be negative for the production efficiency. As to the effect on wage benefit level and new product intensity, both effects don't show the solid significance. On the contrary competition has positive effect on the wage level unsurprisingly. The estimators about the externalities are not that convincing as sample selection can cause bias in estimation.

With respect to each firm's attributes, the production scale has positive effects on the wage benefit level Per capita, new product intensity and patent count. In the perspective of ownership, introduction of foreign, HMT or collective share holders can have positive effect on the production efficiency. And foreign and collective share holders have more propensity to provide higher wages than state-owned share holders. Moreover introduction of foreign share holders have significant propensity to launch new product. As last the stock of invents has significant effect on the wage benefit level and launching new product.

Same as regressing Farrell efficiency on co-patenting network, according to section 8.7 the citation network data are wrangled with serial correlation, scedasticity and cross-section dependence. By means of Hausmman tests, estimators by FE, Arrleo-vcovHC and FEGLS are consistent. Therefore authority have significant positive effect on Farrell efficiency, so firms being heavily citied are more advanced to sharpen their efficiency. On the other side, firms good at absorbing others' patents(Hub) have no significant divergence from others. As to the externalities, urbanization size have positive effect, but the inter-industries diversity have negative effect. Moreover,intra industry diversity or competition level have negative effect. So Marshallian externalities are preferred for improving the Farrell efficiency.

Chapter 5

Conclusions

In chapter 2 the ASIE database is linked with SIPO's and EPO's patent databases respectively with the adoption of data science techniques. Compared with the existing matched results, the linkage between ASIE and SIPO's patent database are multiplied in addition to the improvement of linking precision. Other than firm names, unique organization codes is fully used along the time span, which better cope with the modification of registry info. Furthermore the linguistic obstacles are tamped between ASIE and PATSTAT's citation database. The linked results obviously enrich the perspectives to probe into Chinese manufacture firms' micro performance

In chapter 3 localization quotients and employee weighted KD function are adopted respectively to estimate the degree of agglomeration along the ASIE's time span. The former index is a classical way based on the discrete space presumption. Other than taking a arbitrary value as the threshold, confidence intervals is derived city-sector wise for each year. For the values out of the confidence interval, they are divided into concentration and dispersion groups. On the other side with scraping the geographic information for the ZIP codes in ASIE, each firm is attached with a approximated latitude and longitude. Thereafter I derive KD envelop curves and estimate global localization indices. Based on the estimated results it is found that discrete localization have overestimating bias for agglomeration. Moreover several patterns of geographic distribution evolving progress are summarized.To estimate the relation between firms' size and agglomeration, SAC and SDEM models are adopted to complement the results of basic panel models because cross section dependence exists overwhelmingly. Based on the results of spatial direct and indirect effects, Marshallian externalities are preferred to Jacobian externalities for most sectors' large firms. And the stock of patents have diverging effect. Models' goodness of fit are not solved here as there is no decisive test about the non-nested spatial panel models.

In chapter 4 the co-patenting network is constructed from scratch based on the linking results. According to the visualization and descriptive statistics, the co-patenting network have grown more extensively and sparsely since 1985. Through outcomes of betweenness centrality it is found that the universities in China dominantly occupy central positions in the information flow. On the other side focusing on DOCDB numbers the the citation network centering on Chinese firms are constructed from scratch. According the citation relationship among different regions, United States, Japan, Germany, Taiwan and South Korea maintain the most active interactions with China. Furthermore based on the hub and authority scores(HITS), firms from theses countries overwhelmingly ranks higher than Chinese counterparts in authority, even though Huawei Technology and ZTE uprise in later stage. On the contrary Chinese firms act as the main hubs to absorb technology. In addition, it probes into the effect of co-patenting network's effect on production efficiency, wage level and innovation performance as well as citation network's effect on production efficiency. Especially stochastic frontier models with time-variant efficiency are adopted, through which it is found most sectors experienced convergence of Farrell efficiency and declining marginal effect of incremental invent patents. As to the effect of co-patenting network, Burt's constraint have significantly negative effects on efficiency, wage level and innovation performance. On the contrary betweenness centrality only affects firms' performance indirectly by boosting patent application. In the citation network, firms with higher authority scores boost higher production efficiency. With respect to the effect of agglomeration, sample selection bias probably dampens the derived results. Compared with total number of firms in ASIE, only a small fraction is enlisted in the co-patenting and citation networks. Therefore more appropriate regression model is required to integrate the network and agglomeration.

At last it is ferociously worthy to continue deeper research into the networks and agglomeration's effect on firm's micro level performance. Firstly the citation networks deserve more scrupulous analysis as it is in close relation with firms' technological progress. Secondly firm sorting mechanism is not enrolled here. Firms' location is ordinarily accompanied with the equilibriums of different economic sectors such as land, labor and consumption markets. A more convincing method need to be embedded with broader picture of economy. Thirdly cross dependence exists overwhelming with nearly all regression models about firms. More relaxed econometrics models are needed to cope with this to improve estimators' efficiency. Fourthly micro level network analysis about firms has been emerging these years. Newly developed econometric tools deserves being sharpened for the application to these micro-level databases.

Chapter 6

Appendix to Chapter 2

Table 6.1: Brief statistics of linkage between ASIE and CNKI[26]Part I

Yr	Co-p	Linked	# of Pat	# of Apts	Yr	Co-p	Linked	# of Pat	# of Apts	Yr	Co-p	Linked	# of Pat	# of Apts
1985	F	F	622	226	1990	F	F	809	496	1995	F	F	608	396
		T	171	127			T	286	238			T	316	220
	T	F	226	179		T	F	322	244		T	F	175	144
		T	70	50			T	66	51			T	46	38
1986	F	F	431	266	1991	F	F	1103	625	1996	F	F	1105	589
		T	192	158			T	400	300			T	587	385
	T	F	167	145		T	F	319	230		T	F	447	178
		T	59	52			T	104	70			T	82	52
1987	F	F	489	310	1992	F	F	1223	715	1997	F	F	1098	603
		T	206	171			T	445	351			T	639	424
	T	F	130	119		T	F	326	258		T	F	683	169
		T	51	45			T	107	68			T	99	65
1988	F	F	497	278	1993	F	F	1224	725	1998	F	F	1322	649
		T	166	136			T	462	342			T	754	467
	T	F	202	160		T	F	296	249		T	F	720	197
		T	65	53			T	106	69			T	112	66
1989	F	F	604	318	1994	F	F	1030	606	1999	F	F	2341	752
		T	200	149			T	381	259			T	1329	640
	T	F	170	134		T	F	211	173		T	F	1162	222
		T	54	44			T	59	42			T	273	103

[26]Co-p means whether Co-patenting or not; Linked means whether whether linked to ASIE or not; # of Pat means count of invention patents; # of Apts means count of applicants; T means True; F means False

[27]Ctry stands for the country; sum means current year sum; CumSum stands for cumulative sum; the codes for the country are in accordance with ISO 3166 code

Table 6.2: Brief statistics of linkage between ASIE and CNKI Part II

Yr	Co-p	Linked	# of Pat	# of Apts	Yr	Co-p	Linked	# of Pat	# of Apts	Yr	Co-p	Linked	# of Pat	# of Apts
2000	F	F	8238	1088	2005	F	F	21362	2974	2010	F	F	79415	12711
		T	2168	920			T	18403	3746			T	69005	15654
	T	F	1637	306		T	F	6246	944		T	F	23984	3662
		T	379	131			T	3522	594			T	15932	2694
2001	F	F	4542	1250	2006	F	F	25548	4019	2011	F	F	110993	17574
		T	2962	1084			T	23722	5134			T	95936	20226
	T	F	1770	370		T	F	8087	1173		T	F	29980	4384
		T	380	180			T	4103	781			T	20539	3252
2002	F	F	7927	1773	2007	F	F	34914	5177	2012	F	F	164454	24638
		T	7675	1716			T	34840	6738			T	135236	25439
	T	F	2750	489		T	F	11692	1594		T	F	38359	5489
		T	1050	258			T	7892	1105			T	30796	3857
2003	F	F	10948	2246	2008	F	F	48037	6761	2013	F	F	234440	32031
		T	10011	2425			T	44079	9569			T	158117	27932
	T	F	3543	758		T	F	13131	2181		T	F	51298	6651
		T	1275	379			T	9427	1647			T	46742	4243
2004	F	F	12169	2205	2009	F	F	62007	9882					
		T	13067	2704			T	60526	13911					
	T	F	4000	734		T	F	18045	3076					
		T	1813	435			T	12413	2285					

Table 6.3: Yearly count of Chinese firms in the citation network[27]

Yr	ASIE	Ctry	Sum	CumulSum	Yr	ASIE	Ctry	Sum	CumSum
1988	No	CN	1	1	1989	Yes	CN	1	1
1991	No	CN	2	3	1990	Yes	CN	1	2
1992	No	CN	3	6	1991	Yes	CN	2	4
1994	No	CN	4	10	1993	Yes	CN	2	6
1995	No	CN	7	17	1994	Yes	CN	2	8
1996	No	CN	5	22	1995	Yes	CN	2	10
1997	No	CN	17	39	1996	Yes	CN	6	16
1998	No	CN	32	71	1997	Yes	CN	9	25
1999	No	CN	59	130	1998	Yes	CN	24	49
2000	No	CN	75	205	1999	Yes	CN	40	89
2001	No	CN	123	328	2000	Yes	CN	64	153
2002	No	CN	144	472	2001	Yes	CN	81	234
2003	No	CN	176	648	2002	Yes	CN	124	358
2004	No	CN	207	855	2003	Yes	CN	180	538
2005	No	CN	378	1233	2004	Yes	CN	204	742
2006	No	CN	379	1612	2005	Yes	CN	298	1040
2007	No	CN	427	2039	2006	Yes	CN	371	1411
2008	No	CN	646	2685	2007	Yes	CN	467	1878
2009	No	CN	874	3559	2008	Yes	CN	653	2531
2010	No	CN	1168	4727	2009	Yes	CN	926	3457
2011	No	CN	1205	5932	2010	Yes	CN	1127	4584
2012	No	CN	1622	7554	2011	Yes	CN	1170	5754
2013	No	CN	1748	9302	2012	Yes	CN	1310	7064
					2013	Yes	CN	1318	8382

Table 6.4: Yearly count of Non-Chinese firms in the citation network Part 1

Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum
2004	AE	1	1	2007	BE	7	55	2007	CA	70	268	2013	DD	3	10	2011	ES	38	138
2008	AE	1	2	2008	BE	8	63	2008	CA	62	330	1996	DE	5	5	2012	ES	21	159
2011	AE	1	3	2009	BE	20	83	2009	CA	80	410	1997	DE	8	13	2013	ES	30	189
2012	AE	1	4	2010	BE	18	101	2010	CA	74	484	1998	DE	9	22	1998	FI	1	1
2013	AE	5	9	2011	BE	12	113	2011	CA	97	581	1999	DE	6	28	1999	FI	3	4
2003	AN	2	2	2012	BE	26	139	2012	CA	119	700	2000	DE	18	46	2000	FI	3	7
2006	AN	1	3	2013	BE	21	160	2013	CA	134	834	2001	DE	28	74	2001	FI	4	11
2008	AN	1	4	2007	BG	2	2	2009	CC	1	1	2002	DE	42	116	2002	FI	5	16
2010	AN	1	5	2008	BG	1	3	1998	CH	3	3	2003	DE	53	169	2003	FI	14	30
2005	AR	1	1	2010	BG	1	4	1999	CH	5	8	2004	DE	65	234	2004	FI	10	40
2006	AR	2	3	2011	BG	2	6	2000	CH	6	14	2005	DE	95	329	2005	FI	14	54
2009	AR	2	5	2013	BG	1	7	2001	CH	4	18	2006	DE	144	473	2006	FI	9	63
2011	AR	2	7	1999	BM	1	1	2002	CH	10	28	2007	DE	153	626	2007	FI	7	70
2000	AT	1	1	2000	BM	1	2	2003	CH	18	46	2008	DE	164	790	2008	FI	21	91
2002	AT	2	3	2001	BM	4	6	2004	CH	7	53	2009	DE	202	992	2009	FI	19	110
2003	AT	1	4	2004	BM	1	7	2005	CH	28	81	2010	DE	264	1256	2010	FI	35	145
2004	AT	3	7	2005	BM	3	10	2006	CH	24	105	2011	DE	253	1509	2011	FI	23	168
2005	AT	6	13	2006	BM	3	13	2007	CH	61	166	2012	DE	307	1816	2012	FI	33	201
2006	AT	16	29	2008	BM	4	17	2008	CH	41	207	2013	DE	313	2129	2013	FI	31	232
2007	AT	10	39	2009	BM	1	18	2009	CH	51	258	1996	DK	1	1	1995	FR	1	1
2008	AT	17	56	2010	BM	2	20	2010	CH	64	322	1998	DK	1	2	1997	FR	5	6
2009	AT	20	76	2012	BM	2	22	2011	CH	61	383	1999	DK	2	4	1998	FR	7	13
2010	AT	30	106	2013	BM	5	27	2012	CH	66	449	2000	DK	1	5	1999	FR	7	20
2011	AT	22	128	2005	BN	1	1	2013	CH	67	516	2001	DK	3	8	2000	FR	9	29
2012	AT	25	153	2007	BN	1	2	2010	CK	1	1	2002	DK	1	9	2001	FR	17	46
2013	AT	35	188	2010	BN	1	3	2008	CL	1	1	2003	DK	6	15	2002	FR	10	56
1998	AU	2	2	1999	BR	1	1	2009	CL	5	6	2004	DK	4	19	2003	FR	10	66
1999	AU	6	8	2002	BR	1	2	2010	CL	1	7	2005	DK	9	28	2004	FR	28	94
2000	AU	5	13	2003	BR	1	3	2012	CL	2	9	2006	DK	11	39	2005	FR	25	119
2001	AU	7	20	2004	BR	1	4	2013	CL	3	12	2007	DK	16	55	2006	FR	44	163
2002	AU	2	22	2005	BR	1	5	2008	CO	2	2	2008	DK	15	70	2007	FR	52	215
2003	AU	10	32	2006	BR	3	8	2010	CO	1	3	2009	DK	24	94	2008	FR	58	273
2004	AU	9	41	2007	BR	3	11	2012	CO	3	6	2010	DK	21	115	2009	FR	94	367
2005	AU	18	59	2008	BR	2	13	2013	CO	1	7	2011	DK	23	138	2010	FR	89	456
2006	AU	29	88	2009	BR	3	16	2012	CS	1	1	2012	DK	25	163	2011	FR	104	560
2007	AU	16	104	2010	BR	6	22	2001	CU	1	1	2013	DK	26	189	2012	FR	109	669
2008	AU	24	128	2011	BR	6	28	2006	CU	1	2	2011	EC	1	1	2013	FR	125	794
2009	AU	36	164	2012	BR	6	34	2008	CU	1	3	2008	EE	1	1	1995	GB	3	3
2010	AU	50	214	2013	BR	6	40	2009	CU	1	4	2009	EE	1	2	1997	GB	7	10
2011	AU	35	249	2004	BS	2	2	2003	CW	1	1	2010	EE	2	4	1998	GB	3	13
2012	AU	51	300	2005	BS	2	4	2009	CW	1	2	2011	EE	2	6	1999	GB	10	23
2013	AU	56	356	2006	BS	1	5	2013	CW	1	3	2013	EE	2	8	2000	GB	13	36
2001	BB	1	1	2008	BS	2	7	2004	CY	1	1	2010	EG	1	1	2001	GB	16	52
2002	BB	1	2	2010	BS	1	8	2010	CY	3	4	2012	EG	2	3	2002	GB	30	82
2006	BB	1	3	2013	BS	3	11	2011	CY	3	7	2009	EP	1	1	2003	GB	35	117
2007	BB	2	5	2008	BV	1	1	2013	CY	6	13	2010	EP	3	4	2004	GB	47	164
2008	BB	2	7	2007	BY	1	1	2004	CZ	1	1	2011	EP	2	6	2005	GB	55	219
2012	BB	1	8	2010	BY	1	2	2005	CZ	2	3	2012	EP	1	7	2006	GB	85	304
2013	BB	2	10	2013	BY	2	4	2006	CZ	1	4	2013	EP	7	14	2007	GB	75	379
1988	BE	3	3	2012	BZ	1	1	2007	CZ	4	8	1999	ES	2	2	2008	GB	104	483
1997	BE	1	4	2013	BZ	1	2	2008	CZ	2	10	2001	ES	1	3	2009	GB	110	593
1998	BE	1	5	1998	CA	5	5	2009	CZ	4	14	2002	ES	1	4	2010	GB	120	713
1999	BE	2	7	1999	CA	9	14	2010	CZ	8	22	2003	ES	2	6	2011	GB	157	870
2000	BE	2	9	2000	CA	9	23	2011	CZ	9	31	2004	ES	2	8	2012	GB	205	1075
2001	BE	3	12	2001	CA	14	37	2012	CZ	11	42	2005	ES	6	14	2013	GB	220	1295
2002	BE	2	14	2002	CA	15	52	2013	CZ	9	51	2006	ES	16	30	2008	GC	1	1
2003	BE	5	19	2003	CA	26	78	2006	DD	1	1	2007	ES	16	46	2013	GC	1	2
2004	BE	6	25	2004	CA	22	100	2009	DD	1	2	2008	ES	12	58	2003	GG	1	1
2005	BE	9	34	2005	CA	45	145	2010	DD	2	4	2009	ES	21	79	2011	GN	1	1
2006	BE	14	48	2006	CA	53	198	2011	DD	3	7	2010	ES	21	100	2006	GR	1	1

Table 6.5: Yearly count of Non-Chinese firms in the citation network Part 2

Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum
2007	GR	1	2	2006	IL	35	118	1998	JP	43	92	2012	LI	4	21	2013	MY	2	16
2008	GR	1	3	2007	IL	34	152	1999	JP	41	133	2013	LI	2	23	2010	NI	1	1
2009	GR	2	5	2008	IL	27	179	2000	JP	53	186	2005	LT	1	1	1997	NL	1	1
2010	GR	3	8	2009	IL	42	221	2001	JP	80	266	2008	LT	2	3	1998	NL	2	3
2011	GR	1	9	2010	IL	49	270	2002	JP	98	364	2010	LT	5	8	1999	NL	6	9
2012	GR	4	13	2011	IL	56	326	2003	JP	129	493	2011	LT	2	10	2000	NL	4	13
2013	GR	5	18	2012	IL	86	412	2004	JP	115	608	2012	LT	1	11	2001	NL	5	18
1996	HK	1	1	2013	IL	96	508	2005	JP	228	836	2013	LT	2	13	2002	NL	4	22
1998	HK	3	4	2003	IM	1	1	2006	JP	244	1080	2000	LU	1	1	2003	NL	9	31
1999	HK	1	5	2005	IM	1	2	2007	JP	319	1399	2001	LU	1	2	2004	NL	4	35
2000	HK	5	10	2009	IM	1	3	2008	JP	329	1728	2002	LU	2	4	2005	NL	16	51
2001	HK	6	16	2010	IM	1	4	2009	JP	353	2081	2003	LU	2	6	2006	NL	20	71
2002	HK	7	23	2012	IM	1	5	2010	JP	423	2504	2004	LU	1	7	2007	NL	11	82
2003	HK	3	26	2013	IM	2	7	2011	JP	433	2937	2005	LU	3	10	2008	NL	35	117
2004	HK	9	35	2000	IN	2	2	2012	JP	581	3518	2006	LU	1	11	2009	NL	26	143
2005	HK	24	59	2002	IN	3	5	2013	JP	546	4064	2007	LU	9	20	2010	NL	30	173
2006	HK	20	79	2003	IN	6	11	2004	KN	1	1	2008	LU	4	24	2011	NL	38	211
2007	HK	28	107	2004	IN	5	16	2011	KP	1	1	2009	LU	5	29	2012	NL	50	261
2008	HK	28	135	2005	IN	13	29	1997	KR	1	1	2010	LU	8	37	2013	NL	54	315
2009	HK	40	175	2006	IN	10	39	1998	KR	3	4	2011	LU	13	50	2001	NO	1	1
2010	HK	44	219	2007	IN	11	50	1999	KR	6	10	2012	LU	6	56	2002	NO	5	6
2011	HK	35	254	2008	IN	17	67	2000	KR	11	21	2013	LU	7	63	2003	NO	7	13
2012	HK	48	302	2009	IN	24	91	2001	KR	7	28	2003	LV	1	1	2004	NO	6	19
2013	HK	53	355	2010	IN	18	109	2002	KR	10	38	2009	LV	1	2	2005	NO	4	23
2009	HR	2	2	2011	IN	17	126	2003	KR	32	70	2012	LV	1	3	2006	NO	8	31
2010	HR	1	3	2012	IN	20	146	2004	KR	35	105	2013	LV	2	5	2007	NO	9	40
2012	HR	1	4	2013	IN	29	175	2005	KR	55	160	2010	MC	1	1	2008	NO	8	48
1997	HU	1	1	2005	IR	2	2	2006	KR	62	222	2012	MC	1	2	2009	NO	14	62
2003	HU	3	4	2010	IR	1	3	2007	KR	83	305	2013	MC	2	4	2010	NO	13	75
2006	HU	5	9	2012	IR	1	4	2008	KR	73	378	2006	MO	1	1	2011	NO	18	93
2007	HU	3	12	2001	IS	1	1	2009	KR	101	479	2009	MO	2	3	2012	NO	16	109
2008	HU	1	13	2002	IS	1	2	2010	KR	117	596	2012	MO	2	5	2013	NO	20	129
2009	HU	3	16	2009	IS	2	4	2011	KR	147	743	2013	MO	2	7	2000	NZ	1	1
2010	HU	4	20	2012	IS	2	6	2012	KR	201	944	2008	MT	1	1	2001	NZ	2	3
2011	HU	2	22	1997	IT	3	3	2013	KR	236	1180	2010	MT	2	3	2002	NZ	2	5
2012	HU	6	28	1998	IT	4	7	1997	KY	1	1	2011	MT	4	7	2004	NZ	3	8
2013	HU	5	33	1999	IT	2	9	2002	KY	2	3	2013	MT	3	10	2005	NZ	4	12
1998	IE	1	1	2000	IT	10	19	2003	KY	1	4	2006	MU	1	1	2006	NZ	2	14
1999	IE	1	2	2001	IT	5	24	2004	KY	1	5	2008	MU	3	4	2007	NZ	5	19
2001	IE	2	4	2002	IT	11	35	2005	KY	3	8	2010	MU	1	5	2008	NZ	4	23
2002	IE	1	5	2003	IT	11	46	2006	KY	6	14	2011	MU	2	7	2009	NZ	6	29
2003	IE	7	12	2004	IT	14	60	2007	KY	6	20	2013	MU	1	8	2010	NZ	8	37
2004	IE	6	18	2005	IT	16	76	2008	KY	6	26	2004	MX	1	1	2011	NZ	11	48
2005	IE	5	23	2006	IT	28	104	2009	KY	1	27	2005	MX	2	3	2012	NZ	6	54
2006	IE	4	27	2007	IT	31	135	2010	KY	6	33	2006	MX	2	5	2013	NZ	8	62
2007	IE	8	35	2008	IT	41	176	2011	KY	4	37	2007	MX	2	7	2006	PA	1	1
2008	IE	10	45	2009	IT	41	217	2012	KY	11	48	2008	MX	3	10	2007	PA	1	2
2009	IE	9	54	2010	IT	52	269	2013	KY	13	61	2009	MX	3	13	2013	PA	1	3
2010	IE	20	74	2011	IT	72	341	2013	KZ	1	1	2010	MX	2	15	2010	PG	1	1
2011	IE	18	92	2012	IT	91	432	1997	LI	1	1	2011	MX	2	17	2007	PH	1	1
2012	IE	16	108	2013	IT	81	513	1999	LI	1	2	2012	MX	2	19	2009	PH	2	3
2013	IE	25	133	2008	JM	1	1	2001	LI	1	3	2013	MX	2	21	2010	PH	1	4
1998	IL	1	1	2013	JO	2	2	2002	LI	3	6	2004	MY	2	2	2013	PH	1	5
1999	IL	4	5	1989	JP	1	1	2003	LI	2	8	2005	MY	1	3	2001	PL	1	1
2000	IL	3	8	1991	JP	1	2	2004	LI	1	9	2007	MY	1	4	2003	PL	2	3
2001	IL	9	17	1993	JP	1	3	2005	LI	2	11	2008	MY	1	5	2006	PL	1	4
2002	IL	9	26	1994	JP	1	4	2006	LI	1	12	2009	MY	1	6	2008	PL	1	5
2003	IL	13	39	1995	JP	6	10	2009	LI	2	14	2010	MY	4	10	2009	PL	2	7
2004	IL	20	59	1996	JP	9	19	2010	LI	1	15	2011	MY	1	11	2010	PL	6	13
2005	IL	24	83	1997	JP	30	49	2011	LI	2	17	2012	MY	3	14	2011	PL	1	14

Table 6.6: Yearly count of Non-Chinese firms in the citation network Part 3

Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum	Yr	Ctry	Sum	CSum
2012	PL	11	25	2006	SE	20	68	2012	SU	12	70	2001	US	352	839
2013	PL	10	35	2007	SE	28	96	2013	SU	6	76	2002	US	304	1143
2006	PT	1	1	2008	SE	33	129	2000	TH	1	1	2003	US	384	1527
2007	PT	1	2	2009	SE	40	169	2012	TH	3	4	2004	US	519	2046
2008	PT	1	3	2010	SE	34	203	2013	TH	3	7	2005	US	826	2872
2009	PT	1	4	2011	SE	48	251	2005	TR	1	1	2006	US	750	3622
2010	PT	5	9	2012	SE	70	321	2007	TR	2	3	2007	US	856	4478
2011	PT	9	18	2013	SE	49	370	2008	TR	3	6	2008	US	980	5458
2012	PT	1	19	2000	SG	3	3	2009	TR	6	12	2009	US	1047	6505
2013	PT	7	26	2002	SG	3	6	2010	TR	6	18	2010	US	1228	7733
2013	QA	1	1	2003	SG	7	13	2011	TR	13	31	2011	US	1373	9106
2004	RO	2	2	2004	SG	4	17	2012	TR	9	40	2012	US	1682	10788
2011	RO	1	3	2005	SG	12	29	2013	TR	8	48	2013	US	1999	12787
2013	RO	1	4	2006	SG	21	50	1999	TW	1	1	2013	VC	1	1
2012	RS	1	1	2007	SG	5	55	2000	TW	7	8	2006	VE	1	1
1999	RU	3	3	2008	SG	12	67	2001	TW	9	17	1995	VG	1	1
2000	RU	1	4	2009	SG	13	80	2002	TW	15	32	1998	VG	1	2
2003	RU	2	6	2010	SG	18	98	2003	TW	46	78	2001	VG	5	7
2005	RU	5	11	2011	SG	19	117	2004	TW	48	126	2002	VG	2	9
2006	RU	1	12	2012	SG	20	137	2005	TW	87	213	2003	VG	3	12
2007	RU	16	28	2013	SG	21	158	2006	TW	80	293	2004	VG	4	16
2008	RU	6	34	2004	SI	2	2	2007	TW	96	389	2005	VG	8	24
2009	RU	12	46	2005	SI	2	4	2008	TW	132	521	2006	VG	8	32
2010	RU	18	64	2006	SI	1	5	2009	TW	112	633	2007	VG	8	40
2011	RU	16	80	2007	SI	1	6	2010	TW	181	814	2008	VG	4	44
2012	RU	21	101	2008	SI	1	7	2011	TW	209	1023	2009	VG	1	45
2013	RU	27	128	2009	SI	1	8	2012	TW	179	1202	2010	VG	7	52
2001	SA	1	1	2011	SI	2	10	2013	TW	227	1429	2011	VG	6	58
2003	SA	2	3	2012	SI	4	14	2010	TXKF	1	1	2012	VG	14	72
2004	SA	1	4	2013	SI	2	16	2003	UA	1	1	2013	VG	11	83
2006	SA	1	5	2003	SK	1	1	2005	UA	1	2	2008	WS	2	2
2008	SA	1	6	2008	SK	1	2	2008	UA	1	3	2009	WS	1	3
2009	SA	1	7	2010	SK	1	3	2010	UA	1	4	2011	WS	2	5
2010	SA	2	9	2011	SK	2	5	2012	UA	2	6	2012	WS	1	6
2013	SA	1	10	2012	SK	1	6	2013	UA	1	7	2013	WS	1	7
2011	SC	1	1	2013	SK	2	8	2011	UN	1	1	2001	ZA	1	1
2012	SC	1	2	1998	SU	3	3	1990	US	1	1	2003	ZA	3	4
1995	SE	1	1	1999	SU	1	4	1991	US	1	2	2004	ZA	2	6
1997	SE	2	3	2001	SU	1	5	1992	US	3	5	2005	ZA	2	8
1998	SE	2	5	2002	SU	2	7	1993	US	1	6	2006	ZA	3	11
1999	SE	3	8	2004	SU	2	9	1994	US	6	12	2007	ZA	3	14
2000	SE	2	10	2005	SU	2	11	1995	US	7	19	2008	ZA	3	17
2001	SE	3	13	2007	SU	7	18	1996	US	21	40	2009	ZA	3	20
2002	SE	8	21	2008	SU	2	20	1997	US	48	88	2010	ZA	1	21
2003	SE	7	28	2009	SU	8	28	1998	US	101	189	2011	ZA	3	24
2004	SE	11	39	2010	SU	15	43	1999	US	118	307	2012	ZA	4	28
2005	SE	9	48	2011	SU	15	58	2000	US	180	487	2013	ZA	7	35

Chapter 7

Appendix to Chapter 3

7.1 Manufacture classification codes of GB/T 4754—2002

code	Sector names Chinese	Sector names English
13	农副产品加工业	Processing of food from agric. products
14	食品制造业	Food
15	饮料制造业	Beverage, alcohol and tea product
16	烟草制品业	Tobacco product
17	纺织业	Textile mills
18	纺织服装、服饰业	Apparel, footwear & garment manufacturing
19	皮革、毛皮、羽毛及其制品和制鞋业	leather, fur, feather and related products
20	木材加工和木、竹、藤、棕、草制品业	timber, wood, bamboo, rattan, palm, and straw products
21	家具制造业	Furniture
22	造纸和纸制品业	Paper and paper product
23	印刷和记录媒介复制业	Printing & related activities
24	文教体育用品制造业	Education, art and sports products
25	石油加工、炼焦及核燃料加工业	Petroleum, coking and nuclear fuel process
26	化学原料和化学制品制造业	Chemical raw materials and chemical products
27	医药制造业	Pharmaceutical & medicine product
28	化学纤维制造业	Chemical fibers
29	橡胶制品业	Rubber products
30	塑料制品业	Plastics products
31	非金属矿物制品业	Non-metallic mineral products
32	黑色金属冶炼及压延加工业	Smelting and processing of ferrous metals
33	有色金属冶炼及压延加工业	Smelting and processing of non ferrous metals
34	金属制品业	Metal products
35	通用设备制造业	General purpose machinery
36	专用设备制造业	Special purpose machinery
37	交通运输设备制造业	Transport equipment
39	电气机械及器材制造业	Electrical machinery and equipment
40	通信设备、计算机及其他电子设备制造业	Communication equipment, computers and other electronic equipment
41	仪器仪表及文化、办公用机械制造业	Measuring instruments and machinery for cultural activity and office work
42	工艺品及其他制造业	Artwork and other manufacturing
43	废弃资源和废旧材料回收加工业	Recycling and disposal of waste

Table 7.1: Manufacture classification codes of GB/T 4754—2002

7.2 Mapping LQ values for cities

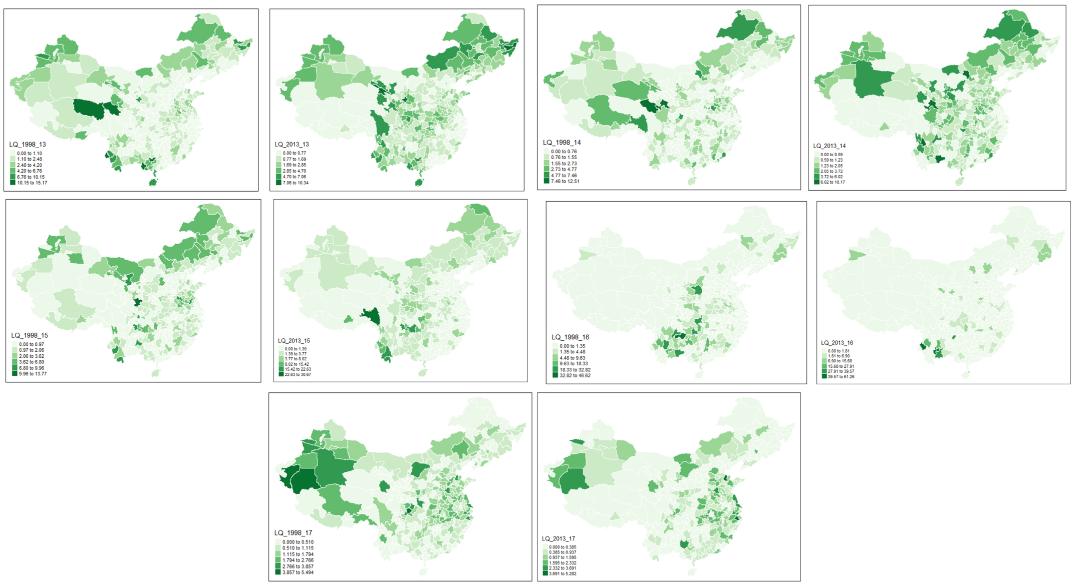


Figure 7.1: Mapping LQ values for cities(sector 13-17)

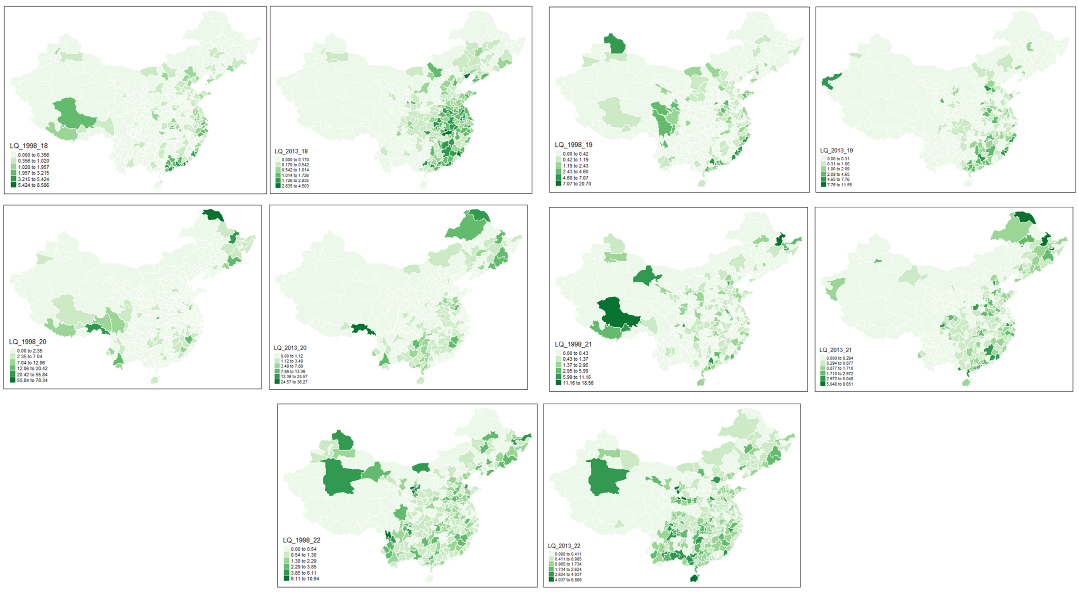


Figure 7.2: Mapping LQ values for cities(sector 18-22)

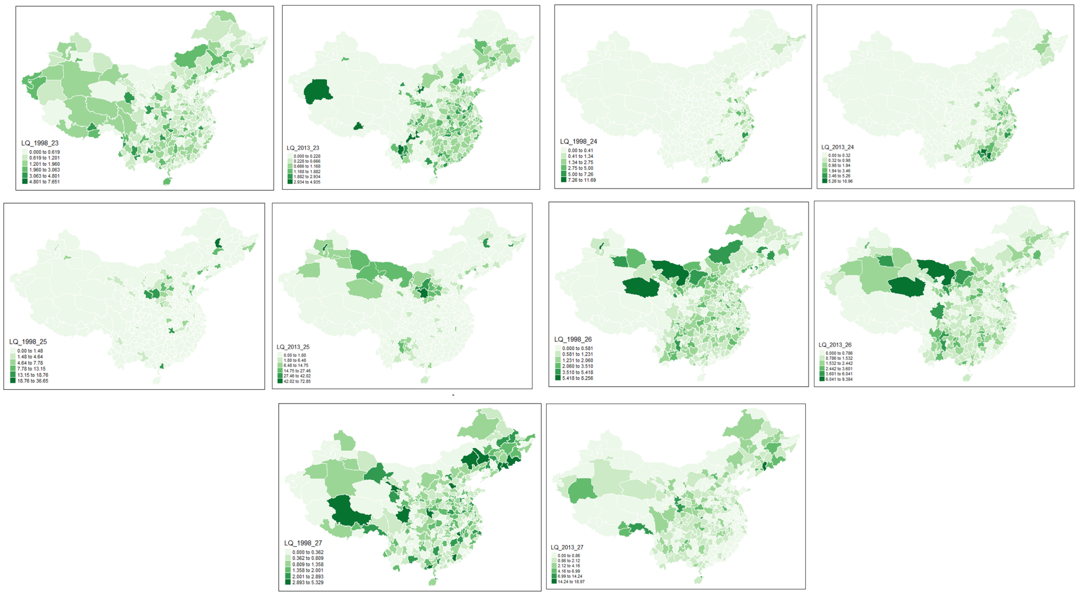


Figure 7.3: Mapping LQ values for cities(sector 23-27)

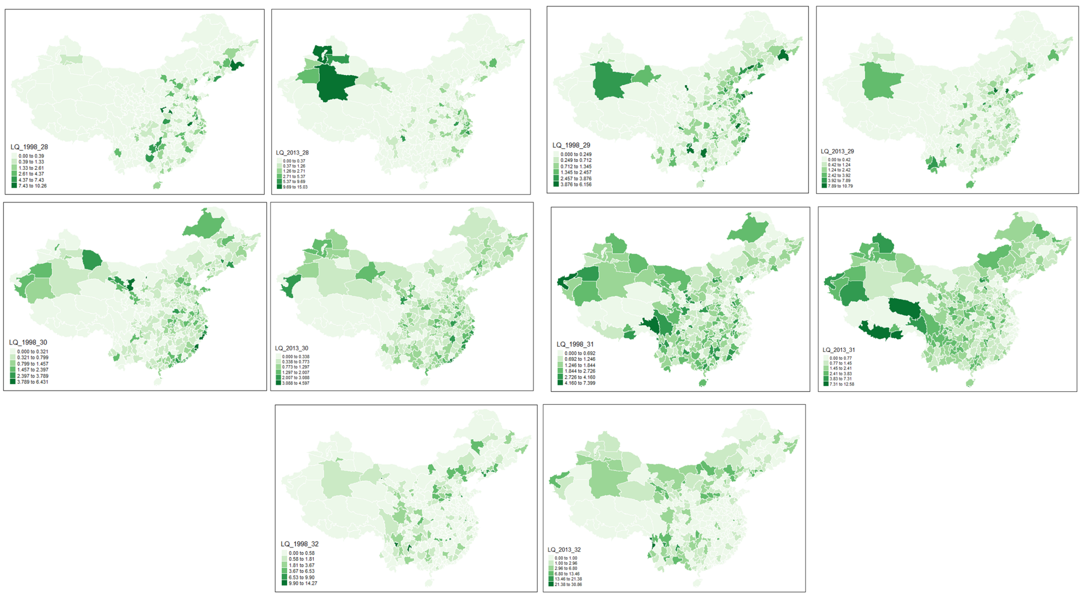


Figure 7.4: Mapping LQ values for cities(sector 28-32)

Figure 7.5: Mapping LQ values for cities(sector 33-37)

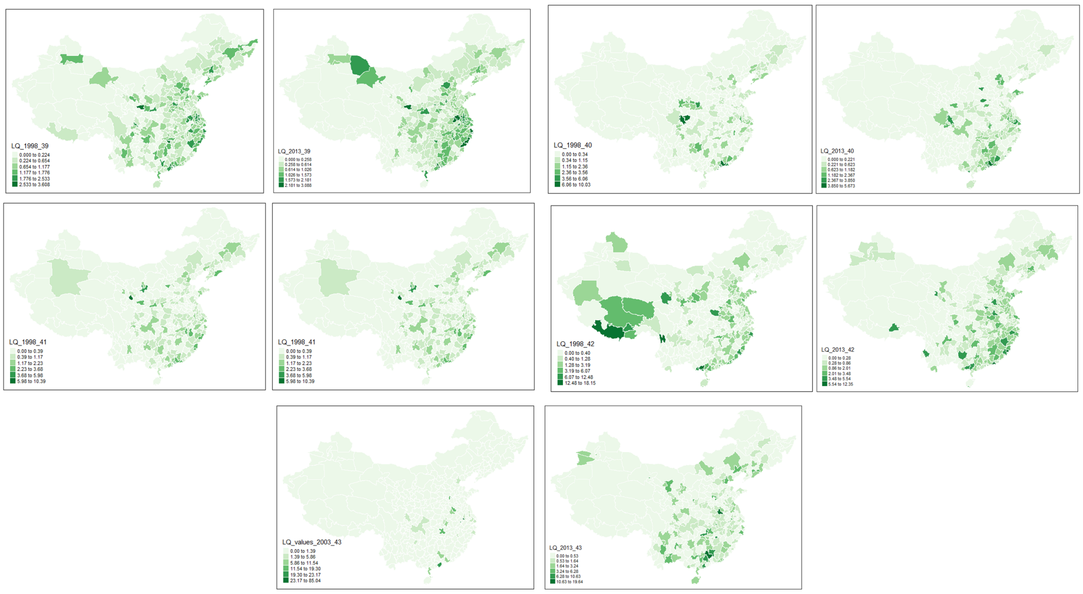


Figure 7.6: Mapping LQ values for cities(sector 39-43)[28]

7.3 Moran' I test result for year 1998, 2006 and 2013

[28]sector 43 was newly enlisted from year 2002

Table 7.2: Moran' I test result for year 1998, 2006 and 2013

sector	year	Moran'I statistic	p-value	sector	year	Moran'I statistic	p-value	sector	year	Moran'I statistic	p-value
13	1998	0.363468041	2.20E-16	23	1998	0.1494	1.03E-05	33	1998	0.059493236	0.05764
13	2006	0.288878309	2.20E-16	23	2006	0.18131	7.90E-13	33	2006	0.160566091	7.85E-07
13	2013	0.34301291	2.20E-16	23	2013	0.1311	9.85E-05	33	2013	0.086721878	5.40E-03
14	1998	0.13045967	9.34E-05	24	1998	0.47596	2.20E-16	34	1998	0.084199399	0.008964
14	2006	0.099671756	0.00231	24	2006	0.46703	2.20E-16	34	2006	0.246060336	4.39E-15
14	2013	0.129128173	0.00012	24	2013	0.44463	2.20E-16	34	2013	0.339935432	2.20E-16
15	1998	0.173092603	2.76E-07	25	1998	0.13419	2.19E-05	35	1998	0.316050044	2.20E-16
15	2006	0.168732209	3.63E-07	25	2006	0.20472	1.56E-10	35	2006	0.342787588	2.20E-16
15	2013	0.28660158	2.20E-16	25	2013	0.26472	2.53E-16	35	2013	0.394319095	2.20E-16
16	1998	0.26708477	2.95E-16	26	1998	0.13785	3.93E-05	36	1998	0.059805583	0.06444
16	2006	0.188080094	5.34E-09	26	2006	0.16273	7.73E-07	36	2006	0.130108177	0.00011
16	2013	0.182974165	5.79E-09	26	2013	0.28577	2.20E-16	36	2013	0.295747048	2.20E-16
17	1998	0.334652962	2.20E-16	27	1998	0.07861	0.01798	37	1998	0.069673216	0.02308
17	2006	0.320889154	2.20E-16	27	2006	0.15826	6.58E-07	37	2006	0.094559639	0.003305
17	2013	0.278396057	2.55E-16	27	2013	0.12194	1.72E-04	37	2013	0.116668767	0.000333
18	1998	0.553951972	2.20E-16	28	1998	0.06925	0.03384	39	1998	0.27654572	5.09E-16
18	2006	0.526676582	2.20E-16	28	2006	0.02195	**0.4511**	39	2006	0.403009419	2.20E-16
18	2013	0.520761955	2.20E-16	28	2013	0.22581	1.60E-12	39	2013	0.390892363	2.20E-16
19	1998	0.339138645	2.20E-16	29	1998	0.10249	0.00211	40	1998	0.314845647	2.20E-16
19	2006	0.219776502	6.40E-13	29	2006	0.05126	**0.1052**	40	2006	0.42430008	2.20E-16
19	2013	0.306140005	2.20E-16	29	2013	0.13486	3.66E-05	40	2013	0.404666249	2.20E-16
20	1998	0.205142279	1.48E-11	30	1998	0.27559	3.94E-16	41	1998	0.167163909	2.20E-07
20	2006	0.225040973	4.78E-12	30	2006	0.26587	8.41E-16	41	2006	0.208758401	4.19E-10
20	2013	0.171478559	4.16E-08	30	2013	0.36259	2.20E-16	41	2013	0.338303683	2.20E-16
21	1998	0.125008551	0.00012	31	1998	0.09491	0.00425	42	1998	0.359225601	2.20E-16
21	2006	0.089619275	0.00275	31	2006	0.17707	1.00E-07	42	2006	0.279389439	2.20E-16
21	2013	0.141870643	1.74E-05	31	2013	0.16899	3.27E-07	42	2013	0.185280821	1.11E-03
22	1998	0.114028366	0.00065	32	1998	0.07647	0.01841	43	2006	0.076428274	0.01604
22	2006	0.052911794	**0.0945**	32	2006	0.13683	4.17E-05	43	2013	0.119396791	1.89E-04
22	2013	0.137338684	4.20E-05	32	2013	0.14265	1.41E-05				

7.4 Global Kernel Density of Each Sector in Year 1998 2003 2008 and 2013 [29]

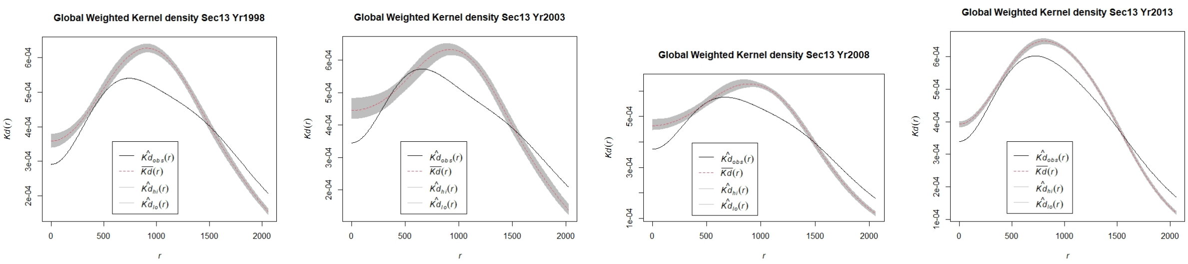


Figure 7.7: sector 13

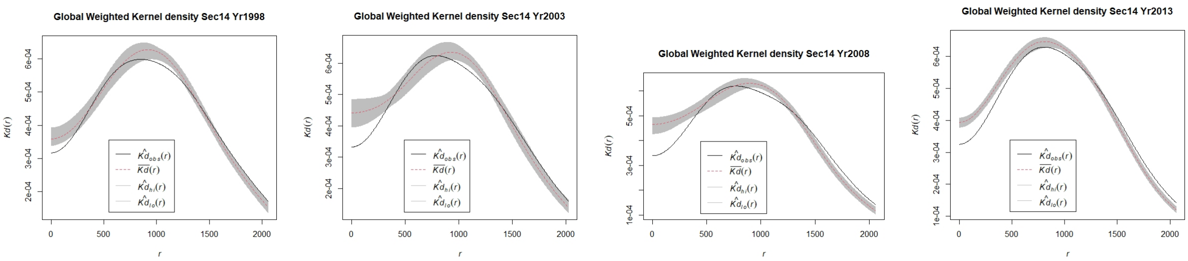


Figure 7.8: sector 14

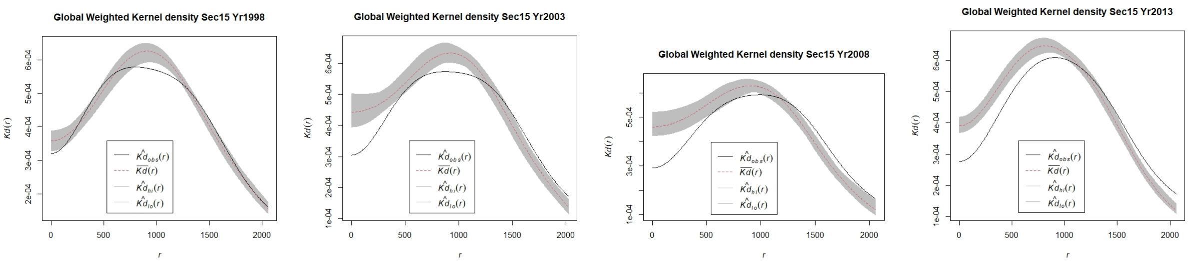


Figure 7.9: sector 15

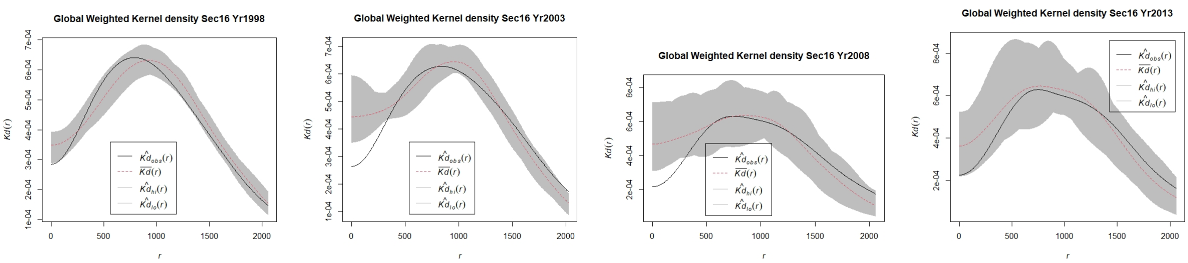


Figure 7.10: sector 16

[29] No firms of sector 33 were enlisted in year 2008

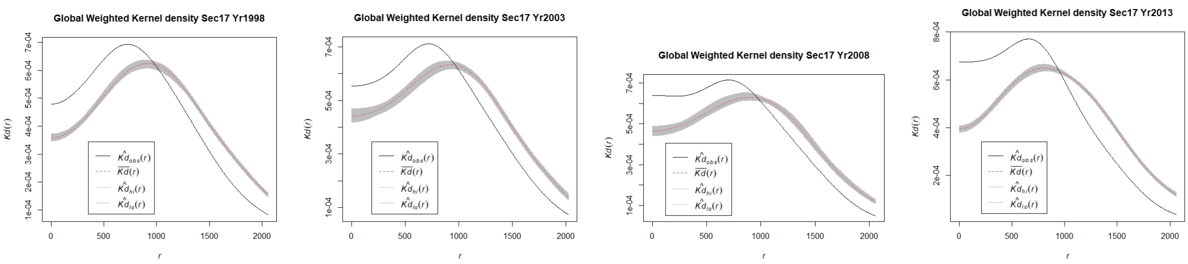


Figure 7.11: sector 17

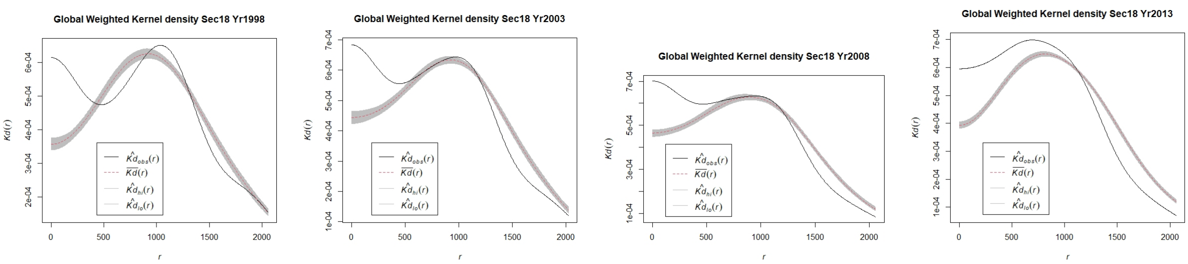


Figure 7.12: sector 18

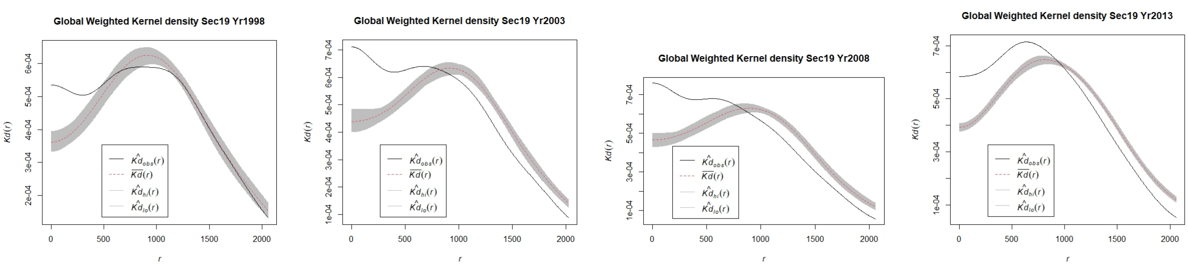


Figure 7.13: sector 19

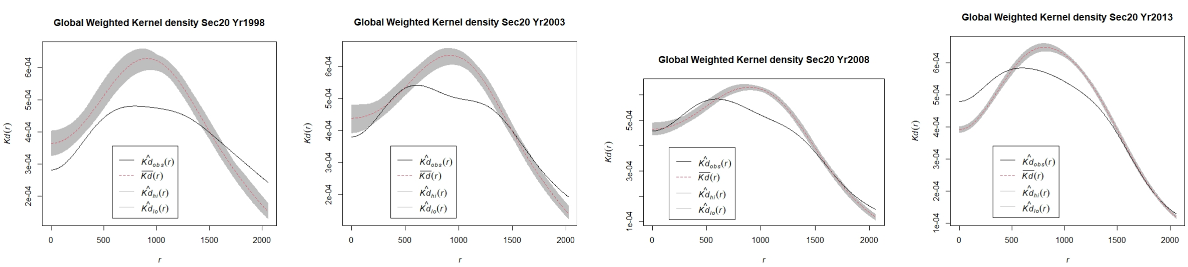


Figure 7.14: sector 20

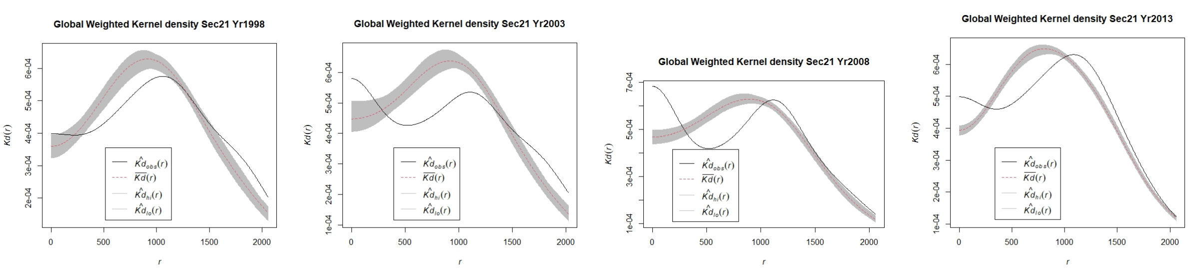
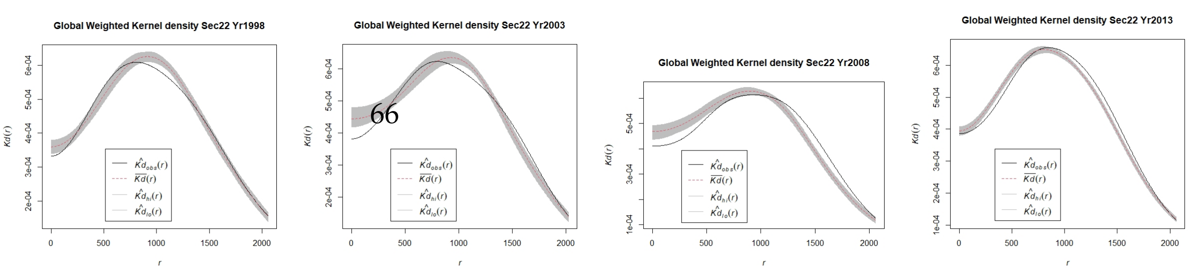


Figure 7.15: sector 21



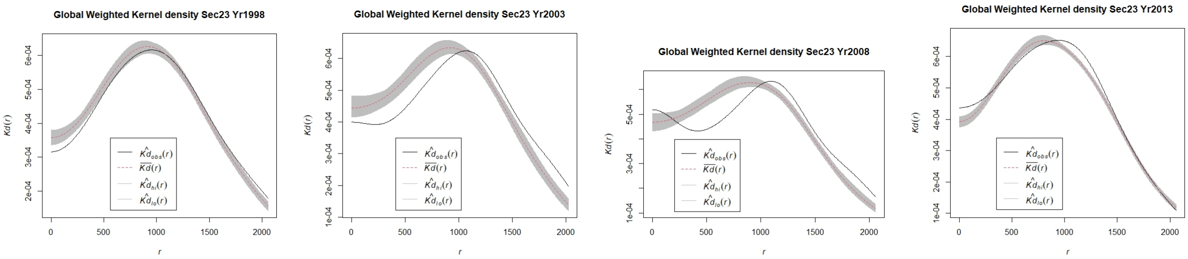


Figure 7.17: sector 23

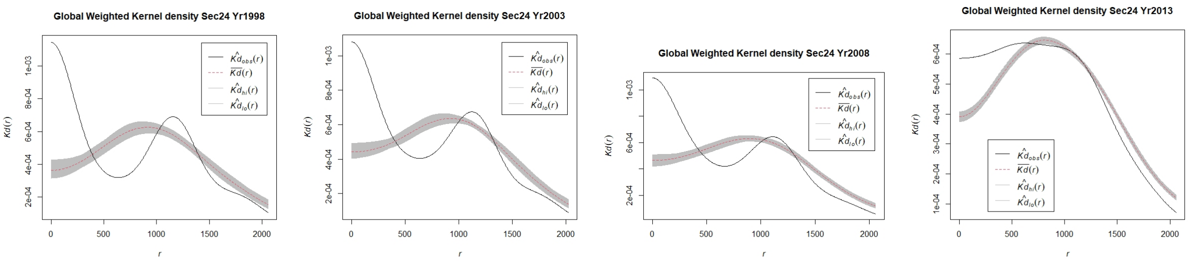


Figure 7.18: sector 24

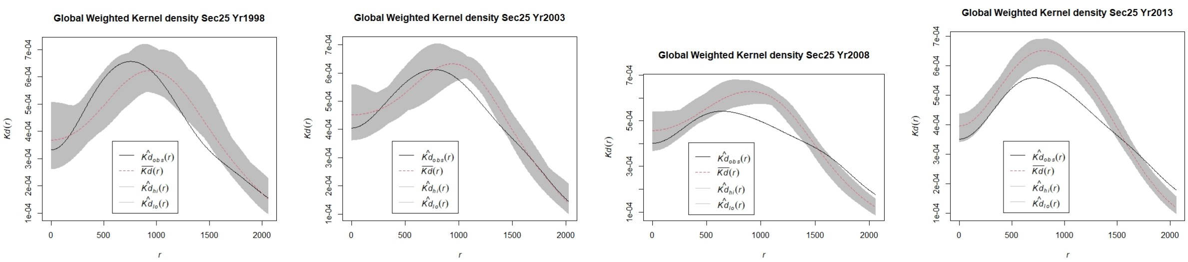


Figure 7.19: sector 25

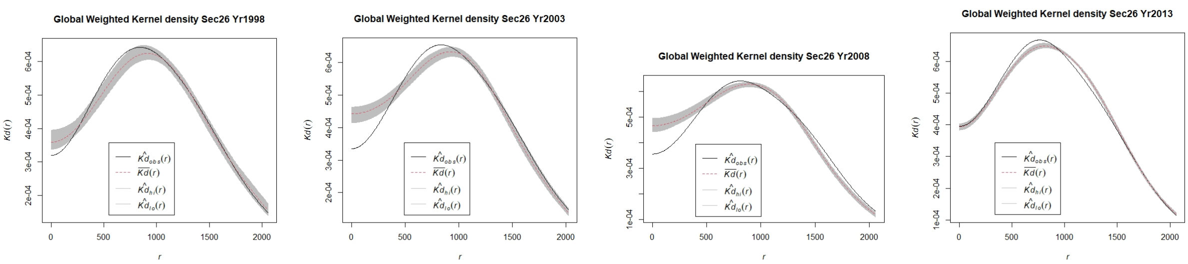


Figure 7.20: sector 26

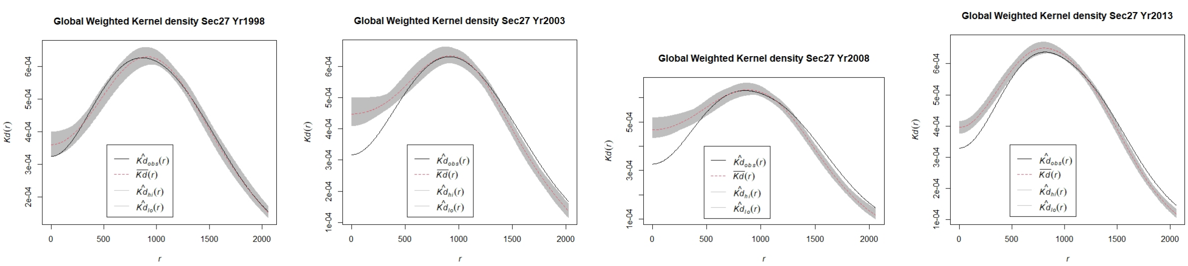


Figure 7.21: sector 27

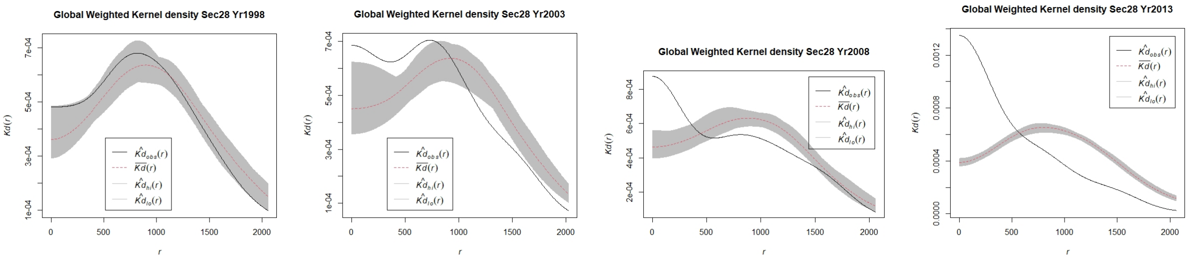


Figure 7.22: sector 28

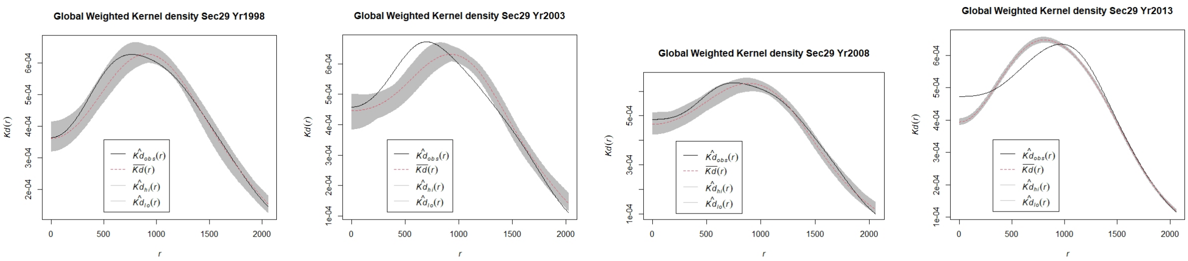


Figure 7.23: sector 29

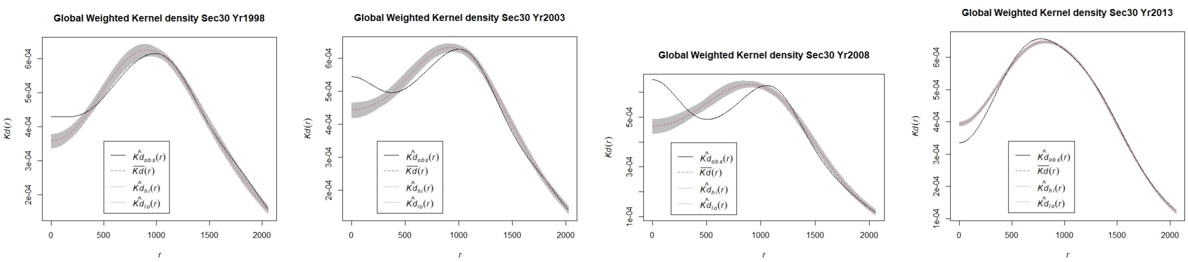


Figure 7.24: sector 30

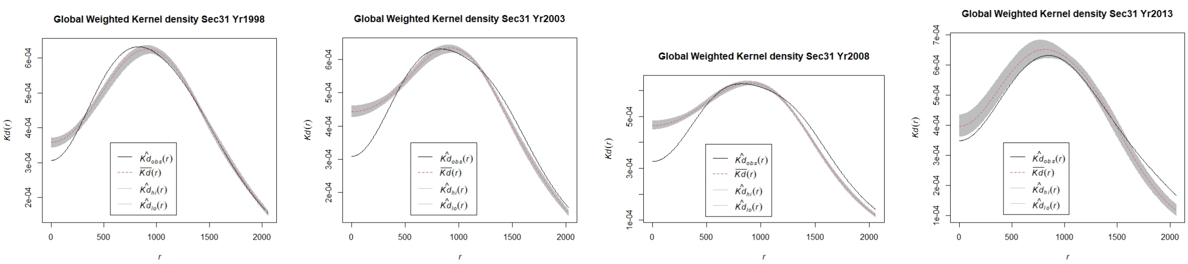


Figure 7.25: sector 31

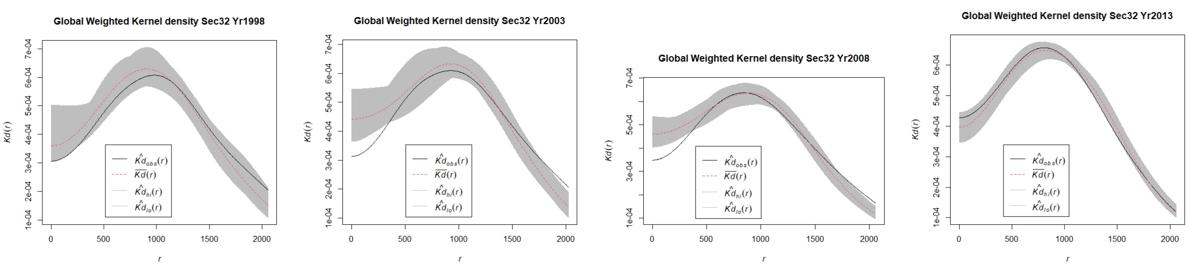


Figure 7.26: sector 32

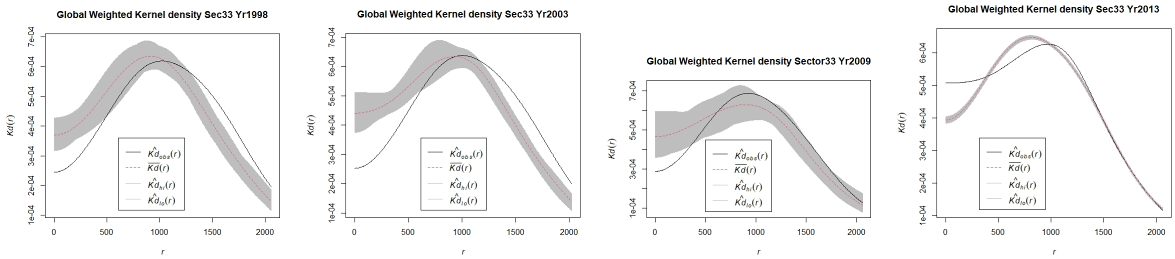


Figure 7.27: sector 33

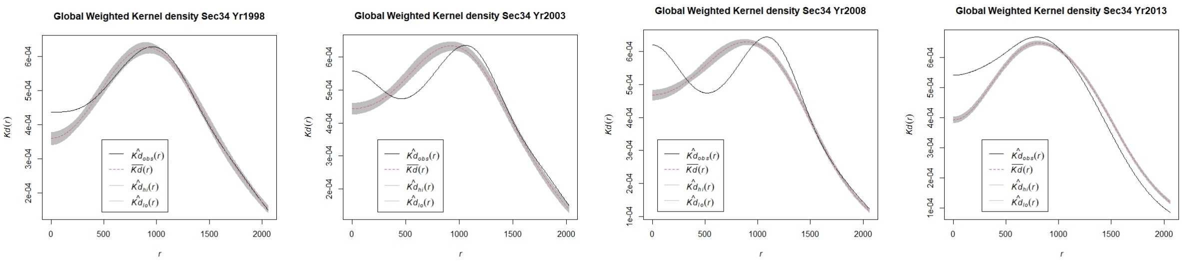


Figure 7.28: sector 34

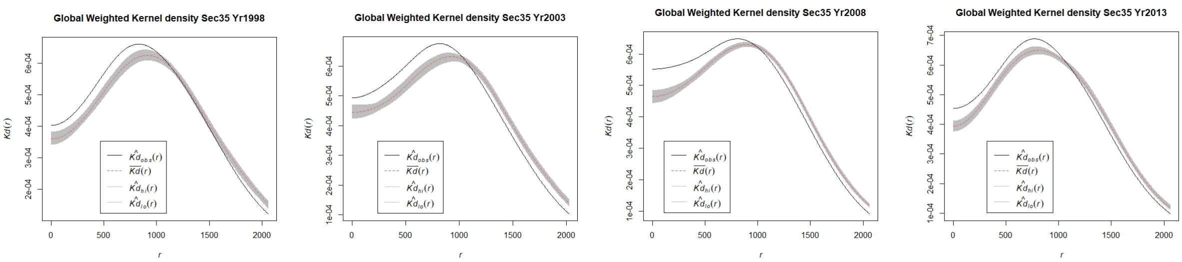


Figure 7.29: sector 35

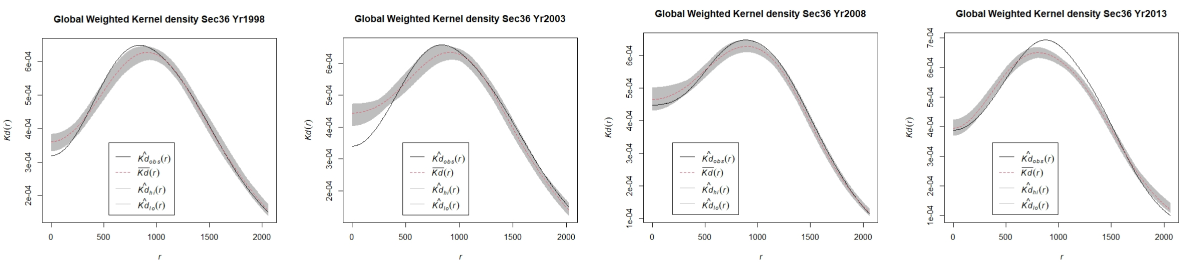


Figure 7.30: sector 36

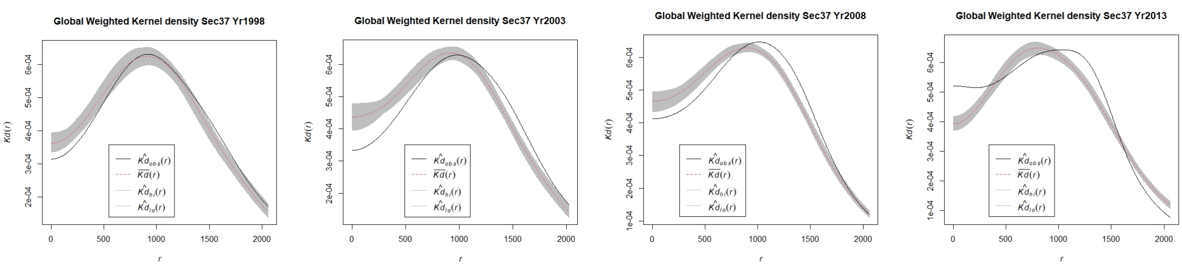


Figure 7.31: sector 37

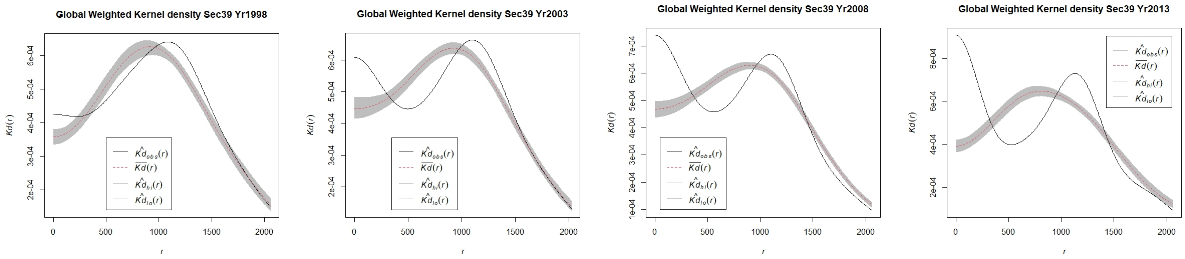


Figure 7.32: sector 39

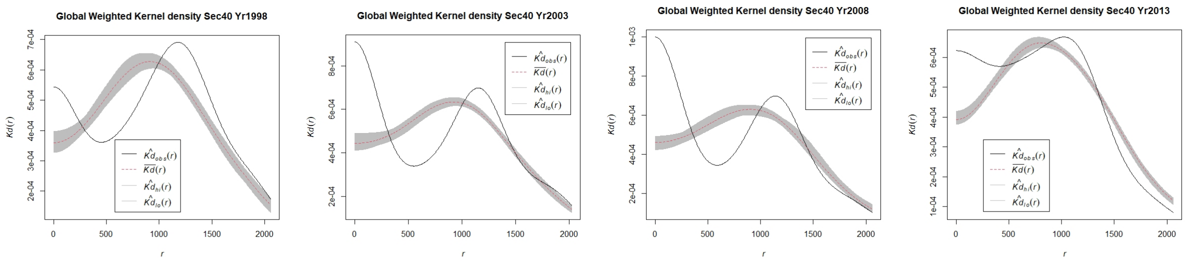


Figure 7.33: sector 40

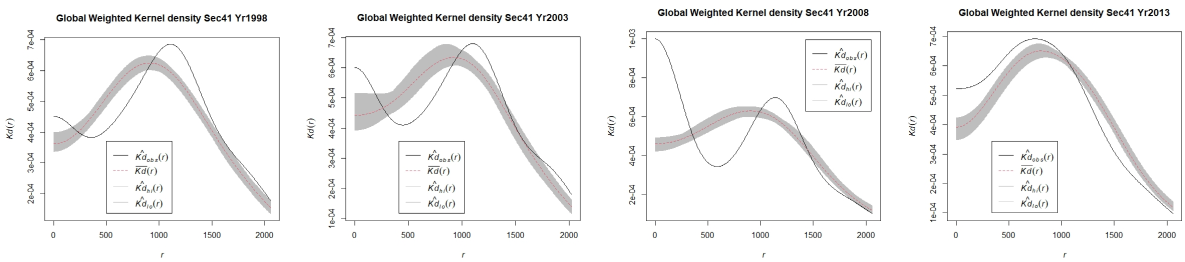


Figure 7.34: sector 41

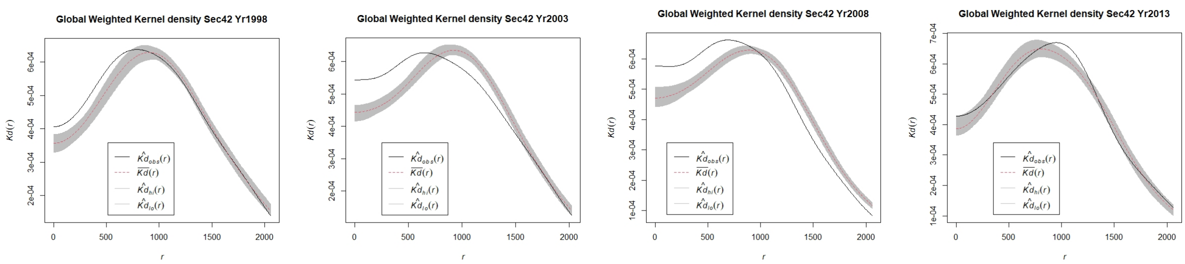


Figure 7.35: sector 42

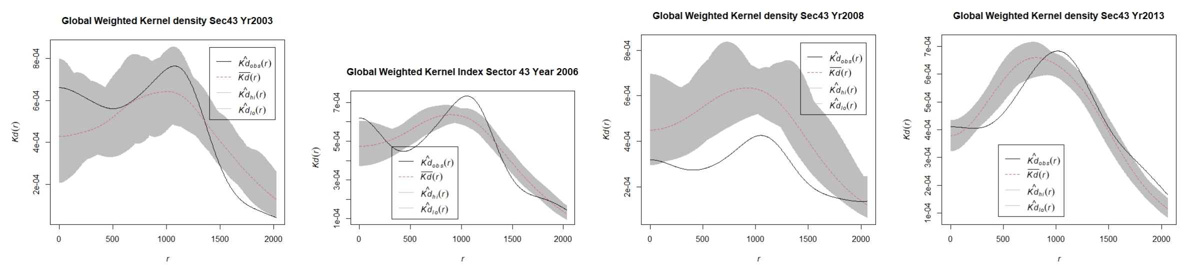


Figure 7.36: sector 43

7.5 Global Localization Shift 1998-2013

7.5.1 Sector 13-27

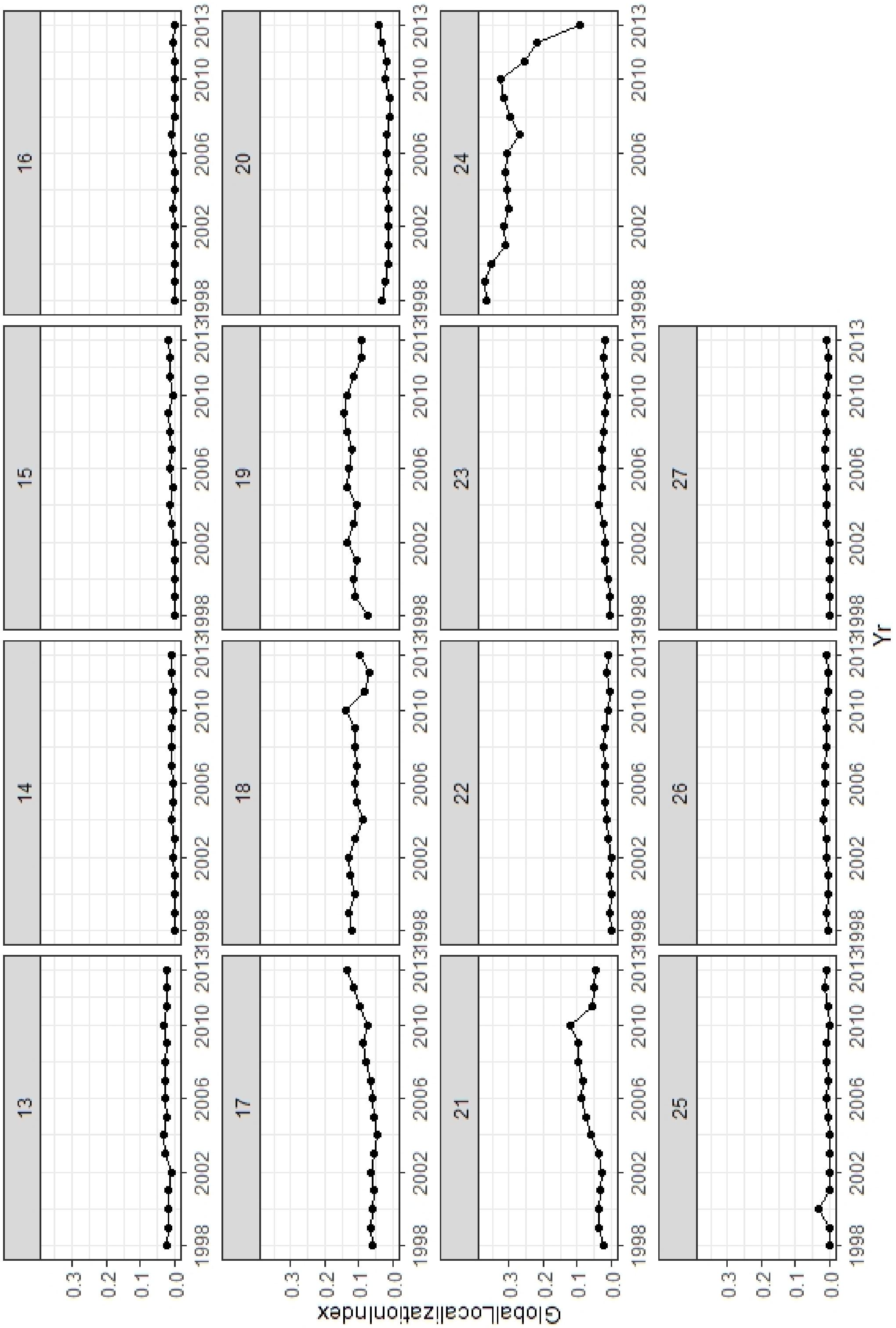


Figure 7.37: Sector 13-27

7.5.2 Sector 28-43

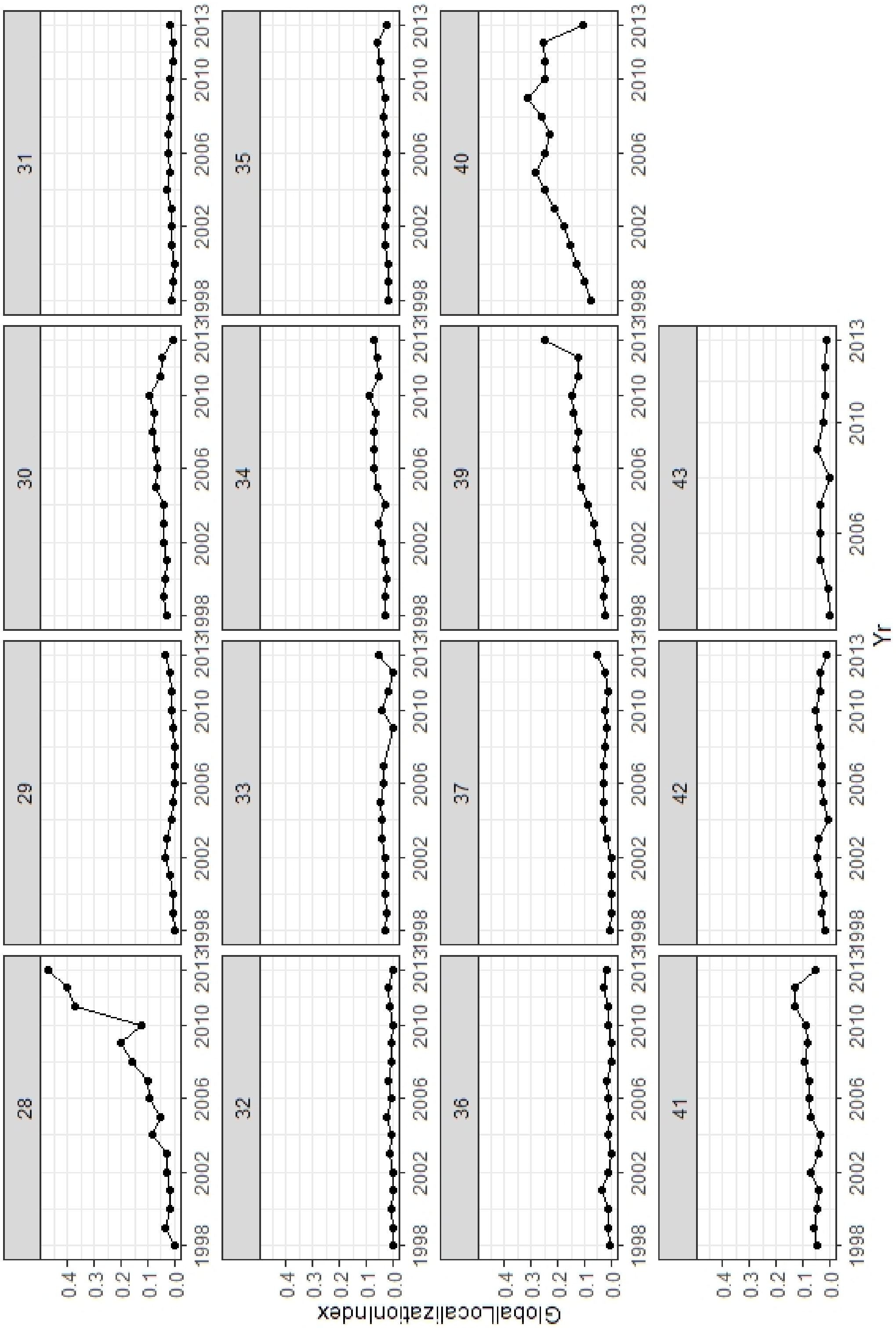


Figure 7.38: Sector 28-43

7.6 Counts of cites grouped by concentration and dispersion for each manufacture sector

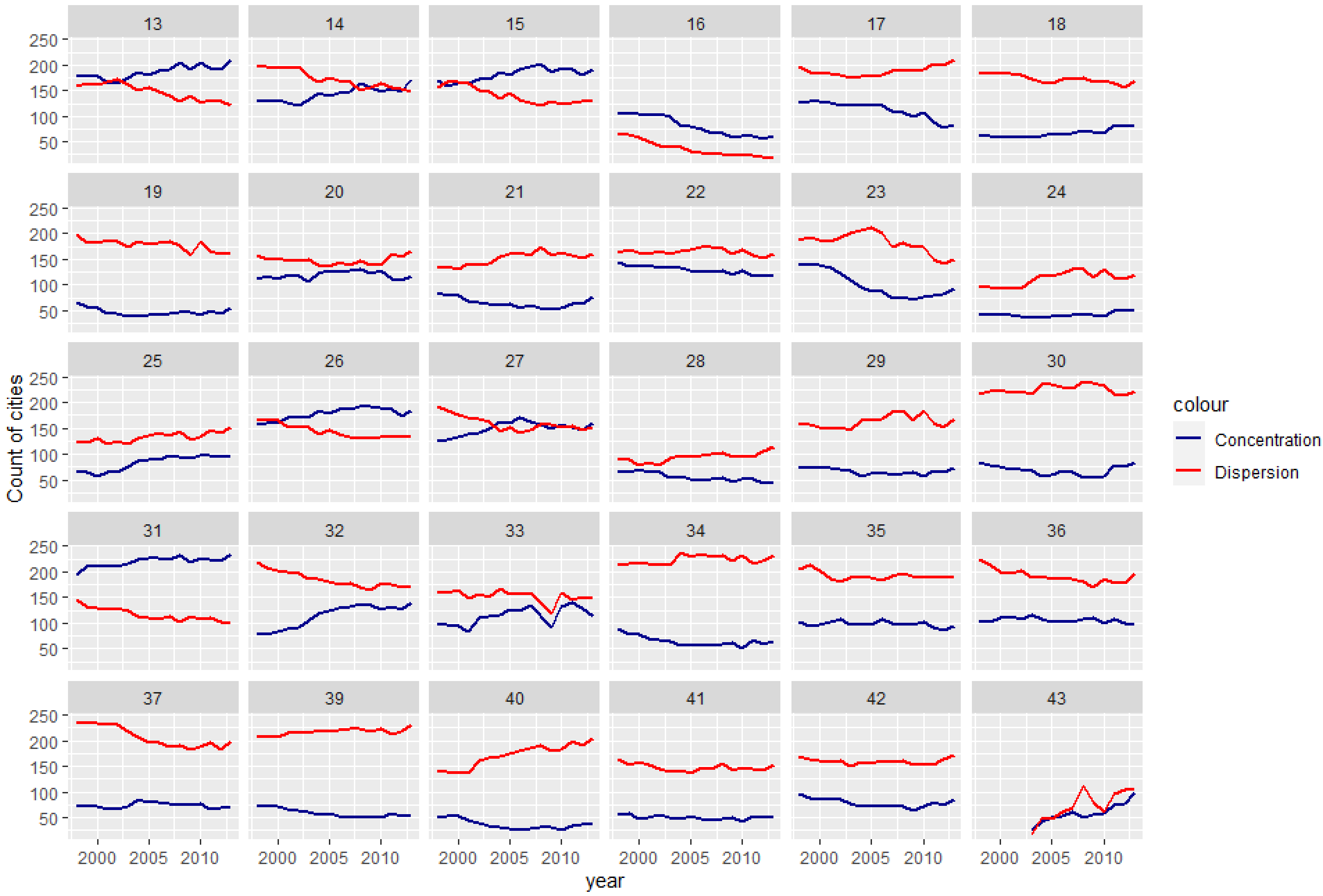


Figure 7.39: Counts of cites grouped by concentration and dispersion for each manufacture sector

7.7 Basic models' regression results for each sector

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.41^{***}		-3.33^{***}		-3.29^{***}
	(0.07)		(0.08)		(0.05)
log(LQ_general)	0.05^{**}	0.15^{***}	0.13^{***}	0.19^{***}	0.12^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.53^{***}	-0.48^{***}	-0.49^{***}	-0.41^{***}	-0.45^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.08^{***}	-0.12^{***}	-0.11^{***}	-0.07^{***}	-0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_ind_E)	0.53^{***}	0.44^{***}	0.45^{***}	0.40^{***}	0.45^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.03	0.06^{***}	0.07^{***}	0.05^{***}	0.04^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.02^{***}	-0.01	-0.01^{*}	-0.01^{**}	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.78	0.69	0.70	0.91	0.77
Adj. R^2	0.78	0.67	0.70		
Num. obs.	5488	5488	5488	5488	5488
s_idios			0.20		
s_id			0.23		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.3: Sector 13

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.07^{***}		-3.66^{***}		-3.63^{***}
	(0.07)		(0.07)		(0.05)
log(LQ_general)	0.09^{***}	0.17^{***}	0.12^{***}	0.22^{***}	0.11^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.64^{***}	-0.61^{***}	-0.62^{***}	-0.57^{***}	-0.58^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.05^{***}	-0.12^{***}	-0.08^{***}	-0.08^{***}	-0.04^{**}
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
log(City_ind_E)	0.56^{***}	0.49^{***}	0.54^{***}	0.45^{***}	0.55^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(City_E)	0.07^{***}	0.04^{*}	0.05^{***}	0.03^{**}	0.03^{**}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.04^{***}	-0.05^{***}	-0.02^{***}	-0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.86	0.77	0.78	0.92	0.85
Adj. R^2	0.86	0.75	0.78		
Num. obs.	5344	5344	5344	5344	5344
s_idios			0.21		
s_id			0.17		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.4: Sector 14

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.97^{***}		-4.51^{***}		-4.14^{***}
	(0.08)		(0.09)		(0.08)
log(LQ_general)	0.25^{***}	0.29^{***}	0.26^{***}	0.20^{***}	0.12^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.67^{***}	-0.63^{***}	-0.64^{***}	-0.59^{***}	-0.62^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.08^{***}	-0.16^{***}	-0.14^{***}	-0.08^{***}	-0.06^{***}
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
log(City_ind_E)	0.38^{***}	0.42^{***}	0.43^{***}	0.52^{***}	0.56^{***}
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
log(City_E)	0.26^{***}	0.16^{***}	0.19^{***}	0.04^{**}	0.07^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.05^{***}	0.01^{*}	-0.00	-0.00	-0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.82	0.69	0.71	0.90	0.80
Adj. R^2	0.82	0.67	0.71		
Num. obs.	5312	5312	5312	5312	5312
s_idios			0.23		
s_id			0.19		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.5: Sector 15

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-5.39^{***}		-5.77^{***}		-5.85^{***}
	(0.07)		(0.10)		(0.09)
log(LQ_general)	-0.22^{***}	-0.28^{***}	-0.22^{***}	-0.10^{***}	-0.11^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
log(1/Competition_HHI)	-1.21^{***}	-1.03^{***}	-1.04^{***}	-0.92^{***}	-0.99^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Diversity_HHI)	-0.01	0.01	-0.00	0.00	-0.01
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	1.15^{***}	1.19^{***}	1.14^{***}	0.99^{***}	1.01^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
log(City_E)	-0.21^{***}	-0.15^{***}	-0.17^{***}	-0.07^{***}	-0.10^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.09^{***}	-0.07^{***}	-0.08^{***}	-0.05^{***}	-0.06^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.96	0.91	0.92	0.97	0.95
Adj. R^2	0.96	0.90	0.92		
Num. obs.	2352	2352	2352	2352	2352
s_idios			0.19		
s_id			0.16		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.6: Sector 16

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.13^{***}		-2.51^{***}		-3.06^{***}
	(0.07)		(0.10)		(0.09)
log(LQ_general)	0.54^{***}	0.57^{***}	0.50^{***}	0.32^{***}	0.28^{***}
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.67^{***}	-0.42^{***}	-0.49^{***}	-0.49^{***}	-0.57^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.15^{***}	-0.25^{***}	-0.22^{***}	-0.14^{***}	-0.12^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.07^{**}	0.01	0.07^{**}	0.33^{***}	0.34^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	0.49^{***}	0.29^{***}	0.32^{***}	0.09^{***}	0.15^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
logCsumYr_pat + 1)	-0.01	0.07^{***}	0.05^{***}	0.04^{***}	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.69	0.49	0.52	0.84	0.65
Adj. R^2	0.69	0.46	0.52		
Num. obs.	5072	5072	5072	5072	5072
s_idios			0.29		
s_id			0.24		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.7: Sector 17

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.14^{***}		-3.20^{***}		-3.54^{***}
	(0.06)		(0.09)		(0.09)
log(LQ_general)	-0.02	0.07^{***}	0.02	0.21^{***}	0.12^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.81^{***}	-0.69^{***}	-0.72^{***}	-0.63^{***}	-0.70^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	0.04^{***}	-0.07^{***}	-0.02	-0.09^{***}	-0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.77^{***}	0.64^{***}	0.69^{***}	0.50^{***}	0.61^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(City_E)	-0.06^{***}	-0.09^{***}	-0.09^{***}	-0.00	-0.00
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.12^{***}	-0.03^{***}	-0.05^{***}	-0.00	-0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.86	0.79	0.80	0.93	0.84
Adj. R^2	0.86	0.78	0.80		
Num. obs.	4224	4224	4224	4224	4224
s_idios			0.22		
s_id			0.20		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.8: Sector 18

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.59^{***}		-3.77^{***}		-4.23^{***}
	(0.07)		(0.08)		(0.07)
log(LQ_general)	0.03	0.14^{***}	0.10^{***}	0.18^{***}	0.10^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.86^{***}	-0.72^{***}	-0.76^{***}	-0.69^{***}	-0.75^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.06^{***}	-0.04^{*}	-0.04^{**}	-0.01	-0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.73^{***}	0.64^{***}	0.67^{***}	0.58^{***}	0.66^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.01	-0.05^{***}	-0.03^{*}	-0.00	0.01
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.03^{***}	0.01	-0.00	0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.90	0.85	0.86	0.95	0.89
Adj. R^2	0.90	0.84	0.86		
Num. obs.	4368	4368	4368	4368	4368
s_idios			0.21		
s_id			0.20		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.9: Sector 19

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.22^{***}		-2.79^{***}		-3.47^{***}
	(0.07)		(0.09)		(0.07)
log(LQ_general)	-0.09^{***}	0.04^{**}	-0.01	0.17^{***}	0.08^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.73^{***}	-0.61^{***}	-0.64^{***}	-0.57^{***}	-0.64^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.05^{***}	-0.10^{***}	-0.08^{***}	-0.06^{**}	-0.04^{*}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.78^{***}	0.66^{***}	0.70^{***}	0.52^{***}	0.61^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	-0.10^{***}	-0.13^{***}	-0.10^{***}	-0.00	0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.05^{***}	0.02^{***}	-0.00	0.00	-0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.80	0.74	0.75	0.89	0.79
Adj. R^2	0.80	0.72	0.75		
Num. obs.	4752	4752	4752	4752	4752
s_idios			0.24		
s_id			0.19		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.10: Sector 20

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-5.03^{***}		-4.15^{***}		-4.31^{***}
	(0.06)		(0.08)		(0.07)
log(LQ_general)	0.20^{***}	0.26^{***}	0.24^{***}	0.27^{***}	0.17^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.87^{***}	-0.77^{***}	-0.80^{***}	-0.74^{***}	-0.79^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.03^{**}	-0.10^{***}	-0.06^{***}	-0.11^{***}	-0.04^{*}
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
log(City_ind_E)	0.58^{***}	0.50^{***}	0.52^{***}	0.49^{***}	0.60^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.18^{***}	0.06^{***}	0.13^{***}	0.05^{***}	0.09^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.07^{***}	0.00	-0.02^{**}	-0.00	-0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.90	0.84	0.85	0.95	0.90
Adj. R^2	0.90	0.83	0.85		
Num. obs.	4112	4112	4112	4112	4112
s_idios			0.19		
s_id			0.14		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.11: Sector 21

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.60^{***}		-2.72^{***}		-2.79^{***}
	(0.09)		(0.09)		(0.08)
log(LQ_general)	0.33^{***}	0.44^{***}	0.39^{***}	0.35^{***}	0.28^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.64^{***}	-0.55^{***}	-0.56^{***}	-0.54^{***}	-0.57^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.16^{***}	-0.25^{***}	-0.22^{***}	-0.17^{***}	-0.15^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.37^{***}	0.28^{***}	0.32^{***}	0.37^{***}	0.43^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
logCity_E)	0.19^{***}	0.14^{***}	0.14^{***}	0.06^{***}	0.06^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.05^{***}	0.02^{**}	0.01	0.01	-0.01^{*}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.82	0.73	0.75	0.91	0.81
Adj. R^2	0.82	0.72	0.75		
Num. obs.	5008	5008	5008	5008	5008
s_idios			0.23		
s_id			0.20		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.12: Sector 22

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.25^{***}		-3.43^{***}		-3.73^{***}
	(0.06)		(0.07)		(0.05)
log(LQ_general)	0.09^{***}	0.29^{***}	0.15^{***}	0.28^{***}	0.10^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.81^{***}	-0.74^{***}	-0.75^{***}	-0.66^{***}	-0.71^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.02	-0.12^{***}	-0.05^{***}	-0.10^{***}	-0.04^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_ind_E)	0.63^{***}	0.39^{***}	0.54^{***}	0.41^{***}	0.60^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.09^{***}	0.06^{***}	0.07^{***}	0.05^{***}	0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.04^{***}	-0.00	-0.03^{***}	-0.01	-0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.85	0.75	0.75	0.93	0.84
Adj. R^2	0.85	0.73	0.75		
Num. obs.	5168	5168	5168	5168	5168
s_idios			0.20		
s_id			0.17		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.13: Sector 23

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.15^{***}		-3.77^{***}		-4.12^{***}
	(0.10)		(0.11)		(0.11)
log(LQ_general)	-0.08^{***}	0.01	-0.01	0.12^{***}	0.06^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.88^{***}	-0.72^{***}	-0.76^{***}	-0.71^{***}	-0.78^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.02	-0.06^{*}	-0.05^{*}	-0.07^{**}	-0.03
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.89^{***}	0.75^{***}	0.77^{***}	0.62^{***}	0.70^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	-0.12^{***}	-0.08^{***}	-0.09^{***}	0.01	-0.02
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.07^{***}	0.00	-0.01	0.01	-0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.91	0.85	0.86	0.95	0.90
Adj. R^2	0.91	0.84	0.86		
Num. obs.	2896	2896	2896	2896	2896
s_idios			0.20		
s_id			0.16		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.14: Sector 24

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-5.61^{***}		-4.72^{***}		-4.51^{***}
	(0.11)		(0.11)		(0.11)
log(LQ_general)	0.36^{***}	0.50^{***}	0.42^{***}	0.37^{***}	0.24^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.85^{***}	-0.64^{***}	-0.66^{***}	-0.63^{***}	-0.69^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.17^{***}	-0.20^{***}	-0.15^{***}	-0.13^{***}	-0.08^{***}
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.42^{***}	0.34^{***}	0.38^{***}	0.43^{***}	0.52^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	0.29^{***}	0.19^{***}	0.21^{***}	0.09^{***}	0.10^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
log(CsumYr_pat + 1)	-0.16^{***}	0.05^{***}	0.01	0.03^{**}	-0.07^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.88	0.84	0.84	0.95	0.87
Adj. R^2	0.88	0.83	0.84		
Num. obs.	4096	4096	4096	4096	4096
s_idios			0.28		
s_id			0.26		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.15: Sector 25

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.44^{***}		-2.67^{***}		-2.71^{***}
	(0.08)		(0.09)		(0.08)
logg(LQ_general)	0.52^{***}	0.51^{***}	0.50^{***}	0.33^{***}	0.31^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.44^{***}	-0.38^{***}	-0.39^{***}	-0.38^{***}	-0.40^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.15^{***}	-0.28^{***}	-0.26^{***}	-0.15^{***}	-0.13^{***}
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
log(City_ind_E)	0.10^{***}	0.15^{***}	0.15^{***}	0.30^{***}	0.31^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
log(City_E)	0.34^{***}	0.22^{***}	0.23^{***}	0.07^{***}	0.10^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
log(CsumYr_pat + 1)	-0.03^{***}	0.03^{***}	0.03^{***}	0.01^{**}	-0.01^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.75	0.62	0.64	0.89	0.73
Adj. R^2	0.75	0.60	0.64		
Num. obs.	5232	5232	5232	5232	5232
s_idios			0.22		
s_id			0.22		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.16: Sector 26

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.98^{***}		-3.63^{***}		-3.54^{***}
	(0.08)		(0.08)		(0.06)
log(LQ_general)	0.09^{***}	0.19^{***}	0.15^{***}	0.20^{***}	0.11^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.61^{***}	-0.55^{***}	-0.57^{***}	-0.52^{***}	-0.55^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.04^{***}	-0.12^{***}	-0.10^{***}	-0.07^{***}	-0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
log(City_ind_E)	0.51^{***}	0.48^{***}	0.50^{***}	0.47^{***}	0.52^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(City_E)	0.08^{***}	0.05^{***}	0.06^{***}	0.03^{**}	0.03^{*}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.02^{***}	0.00	-0.00	-0.00	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.82	0.72	0.73	0.91	0.81
Adj. R^2	0.82	0.70	0.73		
Num. obs.	5168	5168	5168	5168	5168
s_idios			0.19		
s_id			0.17		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.17: Sector 27

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-6.23^{***}		-5.49^{***}		-4.43^{***}
	(0.15)		(0.15)		(0.14)
log(LQ_general)	0.52^{***}	0.52^{***}	0.53^{***}	0.26^{***}	0.23^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
log(1/Competition_HHI)	-0.94^{***}	-0.66^{***}	-0.76^{***}	-0.70^{***}	-0.80^{***}
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.02	-0.14^{***}	-0.07^{*}	-0.08^{*}	-0.03
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
log(City_ind_E)	0.29^{***}	0.26^{***}	0.24^{***}	0.53^{***}	0.57^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
log(City_E)	0.41^{***}	0.36^{***}	0.37^{***}	0.10^{***}	0.09^{***}
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
log(CsumYr_pat + 1)	-0.01	0.08^{***}	0.06^{***}	0.02	-0.02^{*}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.89	0.80	0.83	0.93	0.88
Adj. R^2	0.89	0.79	0.83		
Num. obs.	2912	2912	2912	2912	2912
s_idios			0.31		
s_id			0.18		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.18: Sector 28

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.07^{***}		-3.31^{***}		-3.44^{***}
	(0.10)		(0.11)		(0.09)
log(LQ_general)	0.27^{***}	0.34^{***}	0.33^{***}	0.29^{***}	0.23^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.78^{***}	-0.58^{***}	-0.63^{***}	-0.68^{***}	-0.73^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.12^{***}	-0.21^{***}	-0.19^{***}	-0.12^{***}	-0.09^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.47^{***}	0.41^{***}	0.42^{***}	0.48^{***}	0.54^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(City_E)	0.15^{***}	0.11^{***}	0.12^{***}	0.05^{***}	0.04^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.02^{*}	0.02^{***}	0.02^{*}	0.00	-0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.87	0.79	0.81	0.93	0.86
Adj. R^2	0.87	0.78	0.81		
Num. obs.	4096	4096	4096	4096	4096
s_idios			0.24		
s_id			0.18		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.19: Sector 29

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.13^{***}		-2.43^{***}		-2.84^{***}
	(0.07)		(0.08)		(0.07)
log(LQ_general)	0.03	0.15^{***}	0.10^{***}	0.25^{***}	0.16^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.65^{***}	-0.55^{***}	-0.57^{***}	-0.53^{***}	-0.57^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.07^{***}	-0.14^{***}	-0.11^{***}	-0.10^{***}	-0.07^{***}
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
log(City_ind_E)	0.62^{***}	0.52^{***}	0.56^{***}	0.41^{***}	0.49^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(City_E)	-0.02	-0.06^{***}	-0.04^{**}	0.03^{**}	0.03^{**}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.02^{***}	-0.03^{***}	-0.00	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.79	0.74	0.74	0.88	0.77
Adj. R^2	0.79	0.72	0.74		
Num. obs.	5008	5008	5008	5008	5008
s_idios			0.22		
s_id			0.16		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.20: Sector 30

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.82^{***}		-3.22^{***}		-2.97^{***}
	(0.06)		(0.07)		(0.06)
log(LQ_general)	0.39^{***}	0.41^{***}	0.38^{***}	0.19^{***}	0.14^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
log(1/Competition_HHI)	-0.47^{***}	-0.41^{***}	-0.43^{***}	-0.41^{***}	-0.43^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.10^{***}	-0.26^{***}	-0.20^{***}	-0.14^{***}	-0.08^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_ind_E)	0.20^{***}	0.15^{***}	0.18^{***}	0.32^{***}	0.37^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.31^{***}	0.24^{***}	0.27^{***}	0.07^{***}	0.08^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.03^{***}	0.03^{***}	0.02^{***}	0.02^{***}	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.68	0.51	0.53	0.83	0.66
Adj. R^2	0.68	0.47	0.53		
Num. obs.	5488	5488	5488	5488	5488
s_idios			0.19		
s_id			0.17		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.21: Sector 31

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-5.47^{***}		-4.66^{***}		-4.26^{***}
	(0.08)		(0.10)		(0.10)
log(LQ_general)	0.64^{***}	0.60^{***}	0.59^{***}	0.28^{***}	0.29^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.67^{***}	-0.53^{***}	-0.55^{***}	-0.56^{***}	-0.60^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.08^{***}	-0.22^{***}	-0.18^{***}	-0.12^{***}	-0.08^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.03	0.17^{***}	0.13^{***}	0.44^{***}	0.39^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	0.53^{***}	0.32^{***}	0.39^{***}	0.09^{***}	0.15^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(CsumYr_pat + 1)	-0.11^{***}	0.01	-0.02^{**}	0.02^{*}	-0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.85	0.74	0.76	0.93	0.84
Adj. R^2	0.85	0.73	0.76		
Num. obs.	5168	5168	5168	5168	5168
s_idios			0.30		
s_id			0.25		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.22: Sector 32

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.71^{***}		-3.96^{***}		-3.52^{***}
	(0.08)		(0.10)		(0.08)
log(LQ_general)	0.16^{***}	0.16^{***}	0.18^{***}	0.15^{***}	0.16^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.67^{***}	-0.54^{***}	-0.55^{***}	-0.54^{***}	-0.61^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.15^{***}	-0.16^{***}	-0.18^{***}	-0.08^{**}	-0.08^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.51^{***}	0.57^{***}	0.53^{***}	0.54^{***}	0.51^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.07^{***}	0.09^{***}	0.07^{***}	0.03^{*}	0.04^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.00	0.06^{***}	0.05^{***}	0.04^{***}	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.84	0.77	0.77	0.92	0.83
Adj. R^2	0.84	0.75	0.77		
Num. obs.	4768	4768	4768	4768	4768
s_idios			0.29		
s_id			0.23		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.23: Sector 33

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.42^{***}		-2.56^{***}		-3.06^{***}
	(0.06)		(0.08)		(0.07)
log(LQ_general)	-0.01	0.16^{***}	0.09^{***}	0.28^{***}	0.15^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.69^{***}	-0.55^{***}	-0.59^{***}	-0.55^{***}	-0.61^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.06^{***}	-0.19^{***}	-0.13^{***}	-0.15^{***}	-0.08^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.67^{***}	0.51^{***}	0.57^{***}	0.41^{***}	0.52^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(City_E)	-0.03	-0.05^{**}	-0.04^{**}	0.04^{***}	0.03^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.01^{*}	-0.03^{***}	0.00	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.82	0.73	0.74	0.90	0.80
Adj. R^2	0.82	0.71	0.74		
Num. obs.	5024	5024	5024	5024	5024
s_idios			0.23		
s_id			0.18		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.24: Sector 34

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-2.79^{***}		-2.38^{***}		-2.64^{***}
	(0.10)		(0.11)		(0.08)
log(LQ_general)	0.21^{***}	0.31^{***}	0.29^{***}	0.25^{***}	0.24^{***}
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.57^{***}	-0.44^{***}	-0.47^{***}	-0.49^{***}	-0.53^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.06^{***}	-0.21^{***}	-0.17^{***}	-0.13^{***}	-0.11^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.40^{***}	0.29^{***}	0.30^{***}	0.40^{***}	0.41^{***}
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
log(City_E)	0.08^{**}	0.14^{***}	0.11^{***}	0.07^{***}	0.06^{***}
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	0.00	0.04^{***}	0.04^{***}	0.01	-0.01^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.71	0.55	0.58	0.85	0.70
Adj. R^2	0.71	0.52	0.58		
Num. obs.	4896	4896	4896	4896	4896
s_idios			0.27		
s_id			0.22		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.25: Sector 35

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-5.25^{***}		-4.19^{***}		-3.81^{***}
	(0.09)		(0.10)		(0.10)
log(LQ_general)	0.49^{***}	0.63^{***}	0.57^{***}	0.40^{***}	0.32^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.69^{***}	-0.51^{***}	-0.55^{***}	-0.52^{***}	-0.57^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.08^{***}	-0.21^{***}	-0.17^{***}	-0.10^{***}	-0.06^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.16^{***}	0.04^{*}	0.08^{***}	0.29^{***}	0.37^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	0.47^{***}	0.41^{***}	0.42^{***}	0.17^{***}	0.17^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(CsumYr_pat + 1)	-0.01	0.05^{***}	0.04^{***}	0.02^{***}	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.80	0.65	0.68	0.90	0.78
Adj. R^2	0.80	0.63	0.68		
Num. obs.	5120	5120	5120	5120	5120
s_idios			0.28		
s_id			0.22		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.26: Sector 36

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.62^{***}		-4.16^{***}		-3.68^{***}
	(0.08)		(0.10)		(0.09)
log(LQ_general)	0.13^{***}	0.19^{***}	0.18^{***}	0.20^{***}	0.14^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.69^{***}	-0.54^{***}	-0.57^{***}	-0.52^{***}	-0.56^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.07^{***}	-0.14^{***}	-0.13^{***}	-0.08^{***}	-0.06^{**}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.53^{***}	0.47^{***}	0.47^{***}	0.47^{***}	0.51^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(City_E)	0.13^{***}	0.12^{***}	0.12^{***}	0.04^{***}	0.03^{**}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.01	-0.01^{*}	-0.00	-0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.82	0.68	0.71	0.92	0.81
Adj. R^2	0.82	0.66	0.71		
Num. obs.	4864	4864	4864	4864	4864
s_idios			0.27		
s_id			0.27		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.27: Sector 37

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-3.14^{***}		-2.62^{***}		-2.61^{***}
	(0.07)		(0.09)		(0.09)
log(LQ_general)	-0.04^{*}	0.12^{***}	0.08^{***}	0.27^{***}	0.21^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.64^{***}	-0.54^{***}	-0.56^{***}	-0.49^{***}	-0.53^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.07^{***}	-0.13^{***}	-0.13^{***}	-0.13^{***}	-0.11^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.69^{***}	0.56^{***}	0.58^{***}	0.42^{***}	0.45^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(City_E)	-0.11^{***}	-0.06^{***}	-0.07^{***}	0.03^{**}	0.02
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.03^{***}	-0.03^{***}	-0.01	-0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.83	0.77	0.78	0.92	0.80
Adj. R^2	0.83	0.75	0.78		
Num. obs.	4736	4736	4736	4736	4736
s_idios			0.21		
s_id			0.20		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.28: Sector 39

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.02^{***}		-3.32^{***}		-3.23^{***}
	(0.08)		(0.12)		(0.12)
log(LQ_general)	0.18^{***}	0.29^{***}	0.26^{***}	0.32^{***}	0.25^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.73^{***}	-0.61^{***}	-0.64^{***}	-0.57^{***}	-0.62^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.10^{***}	-0.17^{***}	-0.15^{***}	-0.14^{***}	-0.13^{***}
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
log(City_ind_E)	0.55^{***}	0.45^{***}	0.48^{***}	0.41^{***}	0.46^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
log(City_E)	0.07^{***}	0.06^{***}	0.06^{***}	0.06^{***}	0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.08^{***}	-0.05^{***}	-0.06^{***}	-0.03^{***}	-0.05^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
R^2	0.87	0.83	0.83	0.93	0.86
Adj. R^2	0.87	0.82	0.83		
Num. obs.	3728	3728	3728	3728	3728
s_idios			0.25		
s_id			0.21		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.29: Sector 40

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.58^{***}		-3.67^{***}		-3.37^{***}
	(0.10)		(0.11)		(0.11)
log(LQ_general)	0.30^{***}	0.29^{***}	0.28^{***}	0.26^{***}	0.23^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(1/Competition_HHI)	-0.83^{***}	-0.72^{***}	-0.75^{***}	-0.67^{***}	-0.71^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	-0.04^{**}	-0.14^{***}	-0.11^{***}	-0.09^{***}	-0.06^{**}
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)
log(City_ind_E)	0.47^{***}	0.47^{***}	0.47^{***}	0.49^{***}	0.52^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_E)	0.20^{***}	0.11^{***}	0.12^{***}	0.03^{**}	0.05^{***}
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.06^{***}	-0.02^{***}	-0.03^{***}	-0.02^{**}	-0.04^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
R^2	0.88	0.81	0.82	0.93	0.87
Adj. R^2	0.88	0.80	0.82		
Num. obs.	3712	3712	3712	3712	3712
s_idios			0.24		
s_id			0.19		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.30: Sector 41

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.66^{***}		-3.81^{***}		-4.11^{***}
	(0.07)		(0.08)		(0.07)
log(LQ_general)	0.08^{***}	0.19^{***}	0.10^{***}	0.29^{***}	0.11^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(1/Competition_HHI)	-0.88^{***}	-0.73^{***}	-0.77^{***}	-0.69^{***}	-0.77^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Diversity_HHI)	0.02	-0.10^{***}	-0.04^{*}	-0.09^{***}	-0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
log(City_ind_E)	0.70^{***}	0.57^{***}	0.66^{***}	0.49^{***}	0.66^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log(City_E)	0.06^{***}	0.00	0.01	0.03^{**}	0.03^{**}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
log(CsumYr_pat + 1)	-0.02^{**}	0.02^{***}	0.01	0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.89	0.84	0.85	0.94	0.88
Adj. R^2	0.89	0.83	0.85		
Num. obs.	4512	4512	4512	4512	4512
s_idios			0.23		
s_id			0.18		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.31: Sector 42

	Pooling	FE Model	RE Model	FEGLS	FGLS
(Intercept)	-4.87^{***}		-2.94^{***}		-3.53^{***}
	(0.09)		(0.12)		(0.07)
log(LQ_general)	0.21^{***}	0.22^{***}	0.21^{***}	0.27^{***}	0.20^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(1/Competition_HHI)	-0.97^{***}	-0.82^{***}	-0.83^{***}	-0.81^{***}	-0.85^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
log(1/Diversity_HHI)	-0.02	-0.06^{*}	-0.04	-0.01	0.01
	(0.01)	(0.02)	(0.02)	(0.00)	(0.01)
log(City_ind_E)	0.62^{***}	0.58^{***}	0.58^{***}	0.53^{***}	0.61^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(City_E)	0.15^{***}	-0.07^{***}	0.01	-0.01^{*}	0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
log(CsumYr_pat + 1)	-0.02	0.01	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.90	0.83	0.84	0.95	0.89
Adj. R^2	0.90	0.82	0.84		
Num. obs.	2544	2544	2544	2544	2544
s_idios			0.20		
s_id			0.18		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7.32: Sector 43

7.8 Hypothesis testing

Table 7.33: Serial Correlation Tests for FE and RE models

Breusch-Godfrey/Wooldridge test for serial correlation in panel models for FE models(degree of freedom = 16)					
sector 13	chisq = 1981, p-value < 2.2e-16	sector 23	chisq = 1826.1, p-value < 2.2e-16	sector 33	chisq = 1316.2, p-value < 2.2e-16
sector 14	chisq = 1538.7, p-value < 2.2e-16	sector 24	chisq = 895.63, p-value < 2.2e-16	sector 34	chisq = 1602.5, p-value < 2.2e-16
sector 15	chisq = 1811.1, p-value < 2.2e-16	sector 25	chisq = 1598.2, p-value < 2.2e-16	sector 35	chisq = 2490.4, p-value < 2.2e-16
sector 16	chisq = 710.59, p-value < 2.2e-16	sector 26	chisq = 2041.3, p-value < 2.2e-16	sector 36	chisq = 2186.1, p-value < 2.2e-16
sector 17	chisq = 2377.6, p-value < 2.2e-16	sector 27	chisq = 1621.8, p-value < 2.2e-16	sector 37	chisq = 2133.1, p-value < 2.2e-16
sector 18	chisq = 1294.7, p-value < 2.2e-16	sector 28	chisq = 1280.3, p-value < 2.2e-16	sector 39	chisq = 1514, p-value < 2.2e-16
sector 19	chisq = 1260.3, p-value < 2.2e-16	sector 29	chisq = 1526.5, p-value < 2.2e-16	sector 40	chisq = 1205.8, p-value < 2.2e-16
sector 20	chisq = 1897.3, p-value < 2.2e-16	sector 30	chisq = 1633.5, p-value < 2.2e-16	sector 41	chisq = 1117.9, p-value < 2.2e-16
sector 21	chisq = 1003.2, p-value < 2.2e-16	sector 31	chisq = 2173.6, p-value < 2.2e-16	sector 42	chisq = 1309.6, p-value < 2.2e-16
sector 22	chisq = 1835.3, p-value < 2.2e-16	sector 32	chisq = 1918.3, p-value < 2.2e-16	sector 43	chisq = 967.69, p-value < 2.2e-16

Table 7.34: Exogenousity Test for FE and RE models

Hausman Test of FE and RE Degree of Freedom=6					
sector 13	chisq = 38.596, p-value = 8.588e-07	sector 23	chisq = 3836.5, p-value < 2.2e-16	sector 33	chisq = 589.41, p-value < 2.2e-16
sector 14	chisq = 113.55, p-value < 2.2e-16	sector 24	chisq = 334.94, p-value < 2.2e-16	sector 34	chisq = 252.86, p-value < 2.2e-16
sector 15	chisq = 190.8, p-value < 2.2e-16	sector 25	chisq = 458.91, p-value < 2.2e-16	sector 35	chisq = 1105.1, p-value < 2.2e-16
sector 16	chisq = 85.648, p-value = 2.423e-16	sector 26	chisq = 1433.6, p-value < 2.2e-16	sector 36	chisq = 457.48, p-value < 2.2e-16
sector 17	chisq = 578.46, p-value < 2.2e-16	sector 27	chisq = 281.82, p-value < 2.2e-16	sector 37	chisq = 1048.5, p-value < 2.2e-16
sector 18	chisq = 156.44, p-value < 2.2e-16	sector 28	chisq = 554.13, p-value < 2.2e-16	sector 39	chisq = 169.25, p-value < 2.2e-16
sector 19	chisq = 163.31, p-value < 2.2e-16	sector 29	chisq = 351.93, p-value < 2.2e-16	sector 40	chisq = 94.38, p-value < 2.2e-16
sector 20	chisq = 266.48, p-value < 2.2e-16	sector 30	chisq = 133.16, p-value < 2.2e-16	sector 41	chisq = 108.06, p-value < 2.2e-16
sector 21	chisq = 306.64, p-value < 2.2e-16	sector 31	chisq = 380.58, p-value < 2.2e-16	sector 42	chisq = 234.8, p-value < 2.2e-16
sector 22	chisq = 143.11, p-value < 2.2e-16	sector 32	chisq = 7046.9, p-value < 2.2e-16	sector 43	chisq = 404.88, p-value < 2.2e-16

Table 7.35: Exogenousity Test for FEGLS and FGLS

Hausman Test of FEGLS and FGLS Degree of Freedom=6					
sector 13	chisq = 365.36, p-value < 2.2e-16	sector 23	chisq = 1160.1, p-value < 2.2e-16	sector 33	chisq = 507.01, p-value < 2.2e-16
sector 14	chisq = 72.347, p-value = 1.348e-13	sector 24	chisq = 396.69, p-value < 2.2e-16	sector 34	chisq = 412.73, p-value < 2.2e-16
sector 15	chisq = 465.44, p-value < 2.2e-16	sector 25	chisq = 109.46, p-value < 2.2e-16	sector 35	chisq = 358.26, p-value < 2.2e-16
sector 16	chisq = 1142.4, p-value < 2.2e-16	sector 26	chisq = 1002.5, p-value < 2.2e-16	sector 36	chisq = 1845.2, p-value < 2.2e-16
sector 17	chisq = 1560.6, p-value < 2.2e-16	sector 27	chisq = 252.11, p-value < 2.2e-16	sector 37	chisq = 653.19, p-value < 2.2e-16
sector 18	chisq = 971.65, p-value < 2.2e-16	sector 28	chisq = 433.94, p-value < 2.2e-16	sector 39	chisq = 238.76, p-value < 2.2e-16
sector 19	chisq = 512.36, p-value < 2.2e-16	sector 29	chisq = 193.81, p-value < 2.2e-16	sector 40	chisq = 193.49, p-value < 2.2e-16
sector 20	chisq = 904.42, p-value < 2.2e-16	sector 30	chisq = 236.18, p-value < 2.2e-16	sector 41	chisq = 328.8, p-value < 2.2e-16
sector 21	chisq = 624.26, p-value < 2.2e-16	sector 31	chisq = 1362, p-value < 2.2e-16	sector 42	chisq = 887.2, p-value < 2.2e-16
sector 22	chisq = 398.46, p-value < 2.2e-16	sector 32	chisq = 212.19, p-value < 2.2e-16	sector 43	chisq = 502.39, p-value < 2.2e-16

Table 7.36: Cross section dependence test(Scaled LM) for FE and RE

Scaled LM test for cross-sectional dependence in panels					
sector 13	z = 291.25, p-value < 2.2e-16	sector 23	z = 290.05, p-value < 2.2e-16	sector 33	z = 231.83, p-value < 2.2e-16
sector 14	z = 212.35, p-value < 2.2e-16	sector 24	z = 147.32, p-value < 2.2e-16	sector 34	z = 274.38, p-value < 2.2e-16
sector 15	z = 287.38, p-value < 2.2e-16	sector 25	z = 288.02, p-value < 2.2e-16	sector 35	z = 470.78, p-value < 2.2e-16
sector 16	z = 174.37, p-value < 2.2e-16	sector 26	z = 379.26, p-value < 2.2e-16	sector 36	z = 407.15, p-value < 2.2e-16
sector 17	z = 484.06, p-value < 2.2e-16	sector 27	z = 284.99, p-value < 2.2e-16	sector 37	z = 454.4, p-value < 2.2e-16
sector 18	z = 190.19, p-value < 2.2e-16	sector 28	z = 285.25, p-value < 2.2e-16	sector 39	z = 230.59, p-value < 2.2e-16
sector 19	z = 209.52, p-value < 2.2e-16	sector 29	z = 259.15, p-value < 2.2e-16	sector 40	z = 178.38, p-value < 2.2e-16
sector 20	z = 296.47, p-value < 2.2e-16	sector 30	z = 213.25, p-value < 2.2e-16	sector 41	z = 150.96, p-value < 2.2e-16
sector 21	z = 135.56, p-value < 2.2e-16	sector 31	z = 379.65, p-value < 2.2e-16	sector 42	z = 203.66, p-value < 2.2e-16
sector 22	z = 359.66, p-value < 2.2e-16	sector 32	z = 372.66, p-value < 2.2e-16	sector 43	z = 198.41, p-value < 2.2e-16

Table 7.37: Cross section dependence test(Scaled LM) for FEGLS and FGLS

Scaled LM test for cross-sectional dependence in panels					
sector 13	z = 325.64, p-value < 2.2e-16	sector 23	z = 302.07, p-value < 2.2e-16	sector 33	z = 292.72, p-value < 2.2e-16
sector 14	z = 228.95, p-value < 2.2e-16	sector 24	z = 151.56, p-value < 2.2e-16	sector 34	z = 294.33, p-value < 2.2e-16
sector 15	z = 320.44, p-value < 2.2e-16	sector 25	z = 324.62, p-value < 2.2e-16	sector 35	z = 558, p-value < 2.2e-16
sector 16	z = 244.71, p-value < 2.2e-16	sector 26	z = 446.64, p-value < 2.2e-16	sector 36	z = 577.33, p-value < 2.2e-16
sector 17	z = 669.2, p-value < 2.2e-16	sector 27	z = 296.14, p-value < 2.2e-16	sector 37	z = 485.35, p-value < 2.2e-16
sector 18	z = 211.17, p-value < 2.2e-16	sector 28	z = 403.77, p-value < 2.2e-16	sector 39	z = 254.93, p-value < 2.2e-16
sector 19	z = 217.45, p-value < 2.2e-16	sector 29	z = 336.1, p-value < 2.2e-16	sector 40	z = 180.97, p-value < 2.2e-16
sector 20	z = 353.49, p-value < 2.2e-16	sector 30	z = 228.49, p-value < 2.2e-16	sector 41	z = 165.8, p-value < 2.2e-16
sector 21	z = 137.19, p-value < 2.2e-16	sector 31	z = 454.85, p-value < 2.2e-16	sector 42	z = 221.21, p-value < 2.2e-16
sector 22	z = 379.54, p-value < 2.2e-16	sector 32	z = 516.91, p-value < 2.2e-16	sector 43	z = 211.52, p-value < 2.2e-16

Table 7.38: Cross section dependence test with given spatial weights matrix

Scaled LM test for cross-sectional dependence in panels for FEGLS with given KNN_spatial_weights					
sector 13	z = 111.95, p-value < 2.2e-16	sector 23	z = 101.85, p-value < 2.2e-16	sector 33	z = 146.99, p-value < 2.2e-16
sector 14	z = 69.909, p-value < 2.2e-16	sector 24	z = 52.573, p-value < 2.2e-16	sector 34	z = 121.92, p-value < 2.2e-16
sector 15	z = 105.62, p-value < 2.2e-16	sector 25	z = 115.91, p-value < 2.2e-16	sector 35	z = 223.15, p-value < 2.2e-16
sector 16	z = 70.648, p-value < 2.2e-16	sector 26	z = 186.88, p-value < 2.2e-16	sector 36	z = 220.33, p-value < 2.2e-16
sector 17	z = 220.8, p-value < 2.2e-16	sector 27	z = 96.44, p-value < 2.2e-16	sector 37	z = 235.68, p-value < 2.2e-16
sector 18	z = 142.47, p-value < 2.2e-16	sector 28	z = 162.83, p-value < 2.2e-16	sector 39	z = 146.74, p-value < 2.2e-16
sector 19	z = 65.942, p-value < 2.2e-16	sector 29	z = 118.45, p-value < 2.2e-16	sector 40	z = 137, p-value < 2.2e-16
sector 20	z = 132.14, p-value < 2.2e-16	sector 30	z = 139.31, p-value < 2.2e-16	sector 41	z = 47.43, p-value < 2.2e-16
sector 21	z = 49.691, p-value < 2.2e-16	sector 31	z = 157.89, p-value < 2.2e-16	sector 42	z = 99.344, p-value < 2.2e-16
sector 22	z = 239.4, p-value < 2.2e-16	sector 32	z = 287.91, p-value < 2.2e-16	sector 43	z = 101.47, p-value < 2.2e-16

7.9 Spatial econometric regression results for each sector

Table 7.39: Spatial Autoregressive Combined Model Results

(SAC)Spatial panel fixed effects GM model part 1										
spgm(formula =log(LQ_size) log(LQ_general) + log(1/Competition_HHI) + log(1/Diversity_HHI) + log(City_ind_E) + log(City_E) + log(CsumYr_pat + 1), listw = Knn_neighbor_weights model = "within", lag = T, spatial.error = T, moments = "fullweights", lag.instruments = T, method = "w2sls")										
	Sector 13		Sector 14		Sector 15		Sector 16		Sector 17	
Spatial autoregressive coefficient:										
rho	-0.046	(0.03)	0.046	(0.025)	0.153	(0.026)***	0.153	(0.019)***	0.007	(0.037)
Coefficients:										
log(**LQ_general**)	0.135	(0.018)***	0.1623245	(0.017)***	0.255	(0.016)***	-0.177	(0.019)***	0.387	(0.024)***
log(**1/Competition_HHI**)	-0.47	(0.007)***	-0.608707	(0.007)***	-0.631	(0.009)***	-1.035	(0.021)***	-0.504	(0.009)***
log(**1/Diversity_HHI**)	-0.106	(0.013)***	-0.101463	(0.015)***	-0.151	(0.016)***	0.009	(0.021)	-0.168	(0.019)***
log(**City_ind_E**)	0.459	(0.017)***	0.4982992	(0.017)***	0.439	(0.016)***	1.082	(0.018)***	0.234	(0.023)***
log(**City_E**)	0.053	(0.017)**	0.0319628	(0.015)**	0.141	(0.016)***	-0.109	(0.012)***	0.18	(0.019)***
log(**CsumYr_pat + 1**)	-0.013	(0.004)**	-0.035967	(0.005)**	0.003	(0.006)	-0.06	(0.007)***	0.027	(0.007)***
	Sector 18		Sector 19		Sector 20		Sector 21		Sector 22	
Spatial autoregressive coefficient:										
rho	-0.131	(0.076)	-0.043	(0.019)*	0.005	(0.036)	-0.034	(0.029)	-0.014	(0.068)
Coefficients:										
log(**LQ_general**)	0.138	(0.02)***	0.141	(0.015)***	0.124	(0.017)***	0.252	(0.011)***	0.271	(0.021)***
log(**1/Competition_HHI**)	-0.685	(0.008)***	-0.722	(0.009)***	-0.653	(0.008)***	-0.774	(0.009)***	-0.598	(0.009)***
log(**1/Diversity_HHI**)	-0.077	(0.019)***	-0.034	(0.017)*	-0.075	(0.018)***	-0.076	(0.016)***	-0.198	(0.017)***
log(**City_ind_E**)	0.582	(0.02)***	0.636	(9.016)***	0.591	(0.017)***	0.517	(0.011)***	0.461	(0.02)***
log(**City_E**)	0.005	(0.015)	-0.042	(0.014)**	-0.015	(0.016)	0.05	(0.012)***	0.067	(0.016)
log(**CsumYr_pat + 1**)	-0.012	(0.007)	0.01	(0.008)	0.022	(0.007)**	-0.002	(0.008)	-0.013	(0.006)
	Sector 23		Sector 24		Sector 25		Sector 26		Sector 27	
Spatial autoregressive coefficient:										
rho	-0.106	(0.003)**	0.053468	(0.028)	0.082	(0.02)***	0.27	(0.032)***	-0.073	(0.026)**
Coefficients:										
log(**LQ_general**)	0.244	(0.015)***	0.05	(0.021)*	0.346	(0.024)***	0.383	(0.019)***	0.14	(0.019)***
log(**1/Competition_HHI**)	-0.742	(0.008)***	-0.724	(0.011)***	-0.684	(0.012)***	-0.403	(0.008)***	-0.548	(0.008)***
log(**1/Diversity_HHI**)	-0.101	(0.014)***	-0.067	(0.024)**	-0.145	(0.024)***	-0.237	(0.015)***	-0.09	(0.014)***
log(**City_ind_E**)	0.457	(0.015)***	0.708	(0.021)***	0.49	(0.024)***	0.267	(0.018)***	0.531	(0.019)***
log(**City_E**)	0.068	(0.011)***	-0.056	(0.016)***	0.102	(0.019)***	0.143	(0.016)***	0.023	(0.017)
log(**CsumYr_pat + 1**)	-0.004	(0.007)	0.002	(0.007)	0.04	(0.009)***	0.018	(0.004)***	-0.004	(0.004)
	Sector 28		Sector 29		Sector 30		Sector 31		Sector 32	
Spatial autoregressive coefficient:										
rho	-0.023	(0.029)	-0.039	(0.035)	-0.257	(0.065)***	0.334	(0.031)***	0.6	(0.027)***
Coefficients:										
log(**LQ_general**)	0.297	(0.031)***	0.257	(0.022)***	0.198	(0.02)***	0.327	(0.016)***	0.258	(0.025)***
log(**1/Competition_HHI**)	-0.719	(0.016)***	-0.666	(0.011)***	-0.583	(0.007)***	-0.412	(0.007)***	-0.609	(0.009)***
log(**1/Diversity_HHI**)	-0.089	(0.033)**	-0.164	(0.02)***	-0.112	(0.016)***	-0.217	(0.012)***	-0.14	(0.021)***
log(**City_ind_E**)	0.504	(0.03)***	0.515	(0.022)***	0.496	(0.02)***	0.219	(0.015)***	0.471	(0.025)***
log(**City_E**)	0.152	(0.023)***	0.075	(0.017)***	0.026	(0.016)	0.194	(0.014)***	0.068	(0.021)***
log(**CsumYr_pat + 1**)	0.012	(0.009)	-0.002	(0.007)	-0.02	(0.005)***	0.018	(0.004)***	-0.045	(0.007)***

Table 7.40: Spatial Autoregressive Combined Model Results Part 2

(SAC)Spatial panel fixed effects GM model part 2										
	Sector 33		Sector 34		Sector 35		Sector 36		Sector 37	
Spatial autoregressive coefficient:										
rho	0.198	(0.042)***	-0.144	(0.059)*	0.23	(0.053)***	0.351	(0.044)***	0.52	(0.054)***
Coefficients:										
log(**LQ_general)**	0.146	(0.017)***	0.219	(0.019)***	0.218	(0.024)***	0.368	(0.021)***	0.13	(0.023)***
log(**1/Competition_HHI)**	-0.61	(0.01)***	-0.571	(0.008)***	-0.499	(0.01)***	-0.541	(0.01)***	-0.554	(0.01)***
log(**1/Diversity_HHI)**	-0.126	(0.022)***	-0.163	(0.017)***	-0.15	(0.02)***	-0.161	(0.019)***	-0.103	(0.02)***
log(**City_ind_E)**	0.59	(0.016)***	0.471	(0.018)***	0.43	(0.023)***	0.323	(0.02)***	0.557	(0.022)***
log(**City_E)**	0.074	(0.015)***	0.04	(0.015)**	0.069	(0.021)***	0.21	(0.018)***	0.037	(0.019)*
log(**CsumYr_pat + 1)**	0.012	(0.006)	-0.012	(0.005)*	-0.004	(0.005)	-0.011	(0.005)*	-0.054	(0.005)***
	Sector 39		Sector 40		Sector 41		Sector 42		Sector 43	
Spatial autoregressive coefficient:										
rho	-0.127	(0.093)	-0.043129	(0.038)	-0.087	(0.027)**	-0.127	(0.039)***	-0.263	(0.047)***
Coefficients:										
log(**LQ_general)**	0.184	(0.021)***	0.297	(0.018)***	0.23	(0.023)***	0.237	(0.02)***	0.161	(0.01)***
log(**1/Competition_HHI)**	-0.537	(0.008)***	-0.616	(0.01)***	-0.723	(0.011)***	-0.75	(0.009)***	-0.848	(0.012)***
log(**1/Diversity_HHI)**	-0.132	(0.017)***	-0.157	(0.025)***	-0.118	(0.024)***	-0.086	(0.017)***	-0.008	(0.022)
log(**City_ind_E)**	0.506	(0.02)***	0.451	(0.017)***	0.526	(0.022)***	0.539	(0.019)***	0.661	(0.01)***
log(**City_E)**	0.013	(0.016)	0.089	(0.017)***	0.084	(0.019)***	0.051	(0.015)***	-0.02	(0.013)
log(**CsumYr_pat + 1)**	-0.031	(0.004)***	-0.046	(0.005)***	-0.032	(0.005)***	0.012	(0.007)	0.038	(0.01)***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1										

Table 7.41: Direct and indirect impacts derived from the SAC regression results Part 1

	Sector 13			Sector 14			Sector 15			Sector 16		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log **of**	0.135	-0.0059	0.1291	0.162	0.008	0.17	0.256	0.046	0.302	-0.177	-0.03	-0.207
LQ_general	(0.019)***	(0.004)	(0.019)***	(0.018)***	(0.004)	(0.02)***	(0.016)***	(0.01)***	(0.021)***	(0.019)***	(0.004)***	(0.02)***
log **of**	-0.469	0.02	-0.449	-0.609	-0.03	-0.639	-0.632	-0.113	-0.745	-1.037	-0.185	-1.222
1/Competition	(0.006)***	(0.014)	(0.016)***	(0.007)***	(0.016)	(0.018)***	(0.009)***	(0.023)***	(0.024)***	(0.02)***	(0.028)***	(0.037)***
log **of**	-0.106	0.0046	-0.1014	-0.101	-0.005	-0.106	-0.151	-0.027	-0.178	-0.009	-0.002	-0.011
1/Diversity_HHI	(0.014)***	(0.003)	(0.014)***	(0.016)***	(0.003)	(0.016)***	(0.016)***	(0.006)***	(0.019)***	(0.021)	(0.004)	(0.025)
log **of**	0.459	-0.02	0.439	0.498	-0.024	0.474	0.439	0.079	0.518	1.085	0.193	1.278
City_ind_E	(0.018)***	(0.013)	(0.021)***	(0.018)***	(0.013)	(0.021)***	(0.016)***	(0.017)***	(0.025)***	(0.019)***	(0.027)***	(0.024)***
log **of**	0.0531	-0.0023	0.0508	0.032	0.002	0.034	0.141	0.025	0.166	-0.109	-0.019	-0.128
City_E	(0.018)*	(0.002)	(0.017)*	-0.014	(0.001)	(0.015)	(0.015)***	(0.006)***	(0.019)***	(0.012)***	(0.003)***	(0.013)***
log **of**	-0.0125	0.0005	-0.012	-0.036	-0.002	-0.038	0.002	0.0004	0.0024	-0.06	-0.011	-0.071
CsumYr_pat +1	(0.004)*	(0.0004)	(0.004)*	(0.005)***	(0.001)	(0.005)***	(0.006)	(0.001)	(0.007)	(0.007)***	(0.002)***	(0.008)***
	Sector 17			**Sector 18**			**Sector 19**			**Sector 20**		
log **of**	0.387	0.003	0.389	0.138	-0.016	0.1217	0.141	-0.0058	0.135	0.12396485	3.51E-05	0.124
LQ_general	(0.022)***	(0.015)	(0.026)***	(0.0186)***	(0.008)	(0.01918)***	(0.016)***	(0.0027)	(0.0156)***	(0.0166)***	(0.005)	(0.0188)***
log **of**	-0.504	-0.003	-0.508	-0.685	0.0795	-0.605	-0.7224	0.0298	-0.692	-0.653	-3.00E-03	-0.656
1/Competition	(0.009)***	(0.02)	(0.023)***	(0.008)***	(0.0417)	(0.0417)***	(0.009)***	(0.0136)	(0.0169)***	(0.007)***	(0.024)	(0.026)***
log **of**	-0.168	-0.001	-0.169	-0.077	0.0089	-0.0681	-0.0338	0.0014	-0.032	-0.0748	-2.00E-04	-0.075
1/Diversity_HHI	(0.018)***	(0.006)	(0.018)***	(0.018)***	(0.0052)	(0.0175)***	(0.0151)	(0.0009)	(0.0146)	(0.0164)***	(0.0029)	(0.0167)***
log **of**	0.234	0.002	0.236	0.582	-0.0676	0.5146	0.635	-0.026	0.61	0.591	3.00E-03	0.594
City_ind_E	(0.022)***	(0.009)	(0.025)***	(0.0185)***	(0.0358)	(0.0371)***	(0.016)***	(0.012)	(0.019)***	(0.0162)***	(0.022)	(0.0239)***
log **of**	0.18	0.001	0.181	0.005	-0.0006	0.0047	-0.042	0.0017	-0.04	-0.014919	-7.10E-05	-0.01499
City_E	(0.017)***	(0.007)	(0.02)***	(0.0147)	(0.0019)	(0.0131)	(0.0139)*	(0.0009)	(0.0135)*	(0.015)	(0.0008)	(0.015)
log **of**	0.027	0.0001	0.027	-0.012	0.0014	-0.0107	0.01	-0.0004	0.0098	0.0218	1.50E-04	0.02195
CsumYr_pat +1	(0.007)***	(0.001)	(0.007)***	(0.0078)	(0.0012)	(0.0069)	(0.0074)	(0.0003)	(0.007)	(0.007)**	(0.0008)	(0.007)**
	Sector 21			**Sector 22**			**Sector 23**			**Sector 24**		
log **of**	0.252	-8.40E-03	0.2436	0.2707	-0.0037	0.267	0.2438	-0.0236	0.2202	0.0496	0.0027	0.0525
LQ_general	(0.0097)***	(0.0065)	(0.01147)***	(0.021)***	(0.0192)	(0.029)***	(0.0167)***	(0.0072)***	(0.0171))***	(0.0212)	(0.0022)	(0.0227)
log **of**	-0.774314	2.58E-02	-0.7485	-0.5983	0.0083	-0.59	-0.7422	0.0719	-0.6703	-0.7244	-0.0408	-0.7652
1/Competition	(0.008)***	(0.0202)	(0.0221)***	(0.0095)***	(0.0424)	(0.0442)***	(0.0079)***	(0.0216)***	(0.0228))***	(0.01092)***	(0.0217)	(0.0258)***
log **of**	-0.0756	2.50E-03	-0.0731	-0.19763	0.00273	-0.1949	-0.1011	0.0098	-0.0913	-0.0668	-0.0037	-0.07
1/Diversity_HHI	(0.0166)***	(0.002)	(0.016)***	(0.015)***	(0.0141)	(0.0206)***	(0.0148)***	(0.003)**	(0.014))***	(0.0265)	(0.0028)	(0.0282)
log **of**	0.5173	-1.72E-02	0.5001	0.4613	-0.0063	0.455	0.4573	-0.0443	0.413	0.7085	0.039	0.748
City_ind_E	(0.0103)***	(0.0134)	(0.017)***	(0.0204)***	(0.0328)	(0.0379)***	(0.0158)***	(0.01339)***	(0.0191))***	(0.0218)***	(0.021)	(0.029)***
log **of**	0.0505	-1.70E-03	0.0488	0.0671	-0.0011	0.066	0.0682	-0.0066	0.0616	-0.056	-0.003	-0.059
City_E	(0.0125)***	(0.0015)	(0.0117)***	(0.0163)***	(0.0051)	(0.0175)***	(0.0117)***	(0.0022)**	(0.0109))***	(0.0169)**	(0.0019)	(0.018)**
log **of**	-0.0024872	8.32E-05	-0.002404	-0.0125	0.0002	-0.0123	-0.004	0.0004	-0.0036	0.0023	0.0001	0.002
CsumYr_pat +1	(0.0077)	(0.0003)	(0.007)	(0.00649)	(0.0009)	(0.006)	(0.0059)	(0.0006)	(0.0054)	(0.0077)	(0.0004)	(0.008)
Simlulation Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1												

Table 7.42: Direct and indirect impacts derived from the SAC regression results Part 2

	Sector 25			**Sector 26**			**Sector 27**			**Sector 28**		
	Direct	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**
log **of**	0.3461	0.03	0.376	0.3848	0.1404	0.5252	0.1408	-0.009	0.1311	0.2969	-0.006	0.29
LQ_general	(0.0246)***	(0.0081)***	(0.0275)***	(0.017)***	(0.0218)***	(0.029)***	(0.019)***	(0.00375)*	(0.0177)***	(0.0313)***	(0.007)	(0.033)***
log **of**	-0.684	-0.06	-0.744	-0.404	-0.1474	-0.5514	-0.5485	0.0374	-0.511	-0.7194	0.01619	-0.7032
1/Competition	(0.0124)***	(0.0159)***	(0.0224)***	(0.007)***	(0.023)***	(0.025)***	(0.008)***	(0.013)**	(0.0159)***	(0.016)***	(0.018)	(0.0254)***
log **of**	-0.1451	-0.012	-0.158	-0.238	-0.0868	-0.325	-0.089	0.006	-0.083	-0.089	0.002	-0.087
1/Diversity_HHI	(0.0244)***	(0.0039)**	(0.0265)***	(0.014)***	(0.014)***	(0.023)***	(0.015)***	(0.0025)*	(0.01388)***	(0.0319)**	(0.0026)	(0.031)**
log **of**	0.4901	0.043	0.533	0.267	0.0977	0.365	0.5315	-0.0363	0.4951	0.504	-0.0113	0.493
City_ind_E	(0.0245)***	(0.0116)***	(0.03)***	(0.0168)***	(0.0173)***	(0.0293)***	(0.01868)***	(0.0126)**	(0.0223)***	(0.03)***	(0.0132)	(0.0313)***
log **of**	0.1019	0.009	0.111	0.1434	0.0523	0.1958	0.0233	-0.001	0.0217	0.151	-0.003	0.1483
City_E	(0.0176)***	(0.002)***	(0.019)***	(0.0144)***	(0.0088)***	(0.0197)***	(0.0172)	(0.0013)	(0.016)	(0.02141)***	(0.004)	(0.021)***
log **of**	0.039	0.003	0.043	0.018	0.006	0.0249	-0.004	0.0002	-0.0038	0.012	-0.0002	0.012
CsumYr_pat +1	(0.008)***	(0.0011)**	(0.009)***	(0.0039)***	(0.0016)***	(0.005)***	(0.0033)	(0.0002)	(0.003)	(0.0095)	(0.0004)	(0.009)
	Sector 29			**Sector 30**			**Sector 31**			**Sector 32**		
log **of**	0.2569	-0.0095	0.2474	0.1981	-0.04	0.1575	0.3295	0.1616	0.4911	0.2603	0.384	0.6444
LQ_general	(0.0214)***	(0.0082)	(0.022)***	(0.0211)***	(0.009)***	(0.019)***	(0.0166)***	(0.0237)***	(0.0307)***	(0.02459)***	(0.0445)***	(0.0601)***
log **of**	-0.6663	0.0248	-0.6415	-0.5837	0.1197	-0.4639	-0.4151	-0.2036	-0.6187	-0.615	-0.908	-1.523
1/Competition	(0.0111)***	(0.0209)	(0.0239)***	(0.0075)***	(0.0243)***	(0.025)***	(0.0073)***	(0.0301)***	(0.0322)***	(0.01081)***	(0.1093)***	(0.115)***
log **of**	-0.1644	0.0062	-0.1582	-0.112	0.0229	-0.089	-0.219	-0.1074	-0.3264	-0.1416	-0.2089	-0.35
1/Diversity_HHI	(0.0196)***	(0.0053)	(0.0188)***	(0.0162)***	(0.0053)***	(0.0142)***	(0.011)***	(0.0166)***	(0.0226)***	(0.0231)***	(0.0362)***	(0.0563)***
log **of**	0.5153	-0.0192	0.4961	0.496	-0.1018	0.3946	0.22	0.108	0.328	0.4766	0.7032	1.1798
City_ind_E	(0.0214)***	(0.016)	(0.027)***	(0.0204)***	(0.0212)***	(0.0256)***	(0.016)***	(0.0189)***	(0.0311)***	(0.0238)***	(0.1004)***	(0.118)***
log **of**	0.075	-0.0027	0.0723	0.0258	-0.005	0.0205	0.195	0.095	0.29	0.068	0.1016	0.1706
City_E	(0.0171)***	(0.0026)	(0.0162)***	(0.0169)	(0.0035)	(0.0138)	(0.0151)***	(0.0144)***	(0.0239)***	(0.0196)***	(0.0264)***	(0.0452)***
log **of**	-0.00187	0.00007	-0.0018	-0.02	0.0041	-0.016	0.0182	0.0089	0.0271	-0.0454	-0.0671	-0.1125
CsumYr_pat +1	(0.00666)	(0.0003)	(0.006)	(0.005)***	(0.0013)**	(0.004)***	(0.0038)***	(0.0023)***	(0.0058)***	(0.0071)***	(0.0157)***	(0.0223)***
	Sector 33			**Sector 34**			**Sector 35**			**Sector 36**		
log **of**	0.1459	0.0357	0.1817	0.2188	-0.0276	0.1911	0.219	0.0647	0.2837	0.3704	0.1971	0.5676
LQ_general	(0.0184)***	(0.0106)***	(0.0253)***	(0.0186)***	(0.0096)**	(0.0202)***	(0.0248)***	(0.0224)**	(0.039)	(0.0217)***	(0.0343)***	(0.0453)***
log **of**	-0.6107	-0.1497	-0.7605	-0.5713	0.0722	-0.4991	-0.4998	-0.1478	-0.6477	-0.5445	-0.2897	-0.8342
1/Competition	(0.0091)***	(0.0385)***	(0.0423)***	(0.0075)***	(0.0253)**	(0.0259)***	(0.0104)***	(0.0478)**	(0.051)	(0.0104)***	(0.0507)***	(0.0549)***
log **of**	-0.1263	-0.0309	-0.1572	-0.1636	0.0206	-0.1429	-0.1507	-0.0446	-0.1953	-0.1618	-0.086	-0.2478
1/Diversity_HHI	(0.0233)***	(0.0101)**	(0.0302)***	(0.0161)***	(0.0076)**	(0.0158)***	(0.0199)***	(0.0144)**	(0.028)	(0.0212)***	(0.0193)***	(0.037)***
log **of**	0.5902	0.1447	0.7349	0.4712	-0.0596	0.4116	0.4311	0.1275	0.5586	0.325	0.1729	0.4979
City_ind_E	(0.0174)***	(0.0366)***	(0.0421)***	(0.0174)***	(0.0212)**	(0.024)***	(0.0238)***	(0.0422)**	(0.053)	(0.0209)***	(0.0333)***	(0.0469)***
log **of**	0.0739	0.0181	0.0921	0.0404	-0.0051	0.0353	0.0693	0.0205	0.0898	0.2116	0.1126	0.3243
City_E	(0.0161)***	(0.0059)**	(0.0203)***	(0.0158)*	(0.0025)	(0.0142)*	(0.0222)**	(0.009)*	(0.0294)	(0.0178)***	(0.0201)***	(0.0314)***
log **of**	0.0118	0.0028	0.014	-0.0115	0.0014	-0.0101	-0.0036	-0.001	-0.0047	-0.0107	-0.0057	-0.0165
CsumYr_pat +1	(0.0065)	(0.0016)	(0.008)	(0.0049)*	(0.0008)	(0.0044)*	(0.005)	(0.0016)	(0.0066)	(0.0048)*	(0.0029)*	(0.0076)*
Simlulation Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1												

Table 7.43: Direct and indirect impacts derived from the SAC regression results Part 3

	Sector 37			Sector 39			Sector 40		
	Direct	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**
log **of**	0.1315	0.1385	0.27	0.1843	-0.0207	0.1636	0.2968	-0.0122	0.2845
LQ_general	(0.023)***	(0.0437)**	(0.062)***	(0.0213)***	(0.0137)	(0.0251)***	(0.0152)***	(0.01)	(0.0175)***
log **of**	-0.562	-0.5917	-1.1538	-0.5371	0.0604	-0.4767	-0.616	0.0254	-0.5905
1/Competition	(0.0096)***	(0.1307)***	(0.1329)***	(0.00797)***	(0.04)	(0.0407)***	(0.0099)***	(0.0205)	(0.0235)***
log **of**	-0.1043	-0.1098	-0.2141	-0.1321	0.0148	-0.11723	-0.1565	0.0064	-0.15
1/Diversity_HHI	(0.0202)***	(0.0349)**	(0.0511)***	(0.0157)***	(0.0101)	(0.017)***	(0.025)***	(0.0052)	(0.0248)***
log **of**	0.5647	0.5945	1.1592	0.506	-0.0569	0.449	0.4513	-0.0186	0.4326
City_ind_E	(0.021)***	(0.1302)***	(0.1343)***	(0.0198)***	(0.0379)	(0.0393)***	(0.0151)***	(0.015)	(0.0214)***
log **of**	0.0375	0.039	0.077	0.0125	-0.0014	0.0111	0.0894	-0.0037	0.0857
City_E	(0.0172)	(0.0205)*	(0.0367)*	(0.0148)	(0.0021)	(0.013)	(0.0162)***	(0.003)	(0.016)***
log **of**	-0.0551	-0.058	-0.1132	-0.0309	0.0034	-0.0274	-0.0458	0.0018	-0.044
CsumYr_pat +1	(0.0048)***	(0.0148)***	(0.0179)***	(0.0045)***	(0.0023)	(0.0046)***	(0.0049)***	(0.0015)	(0.00498)***
	Sector 41			**Sector 42**			**Sector 43**		
log **of**	0.2309	-0.0185	0.2123	0.2367	-0.026	0.2098	0.162	-0.034	0.128
LQ_general	(0.0235)***	(0.0051)***	(0.0231)***	(0.0179)***	(0.0067)***	(0.017)***	(0.0097)***	(0.0051)***	(0.0098)***
log **of**	-0.7231	0.058	-0.665	-0.7502	0.085	-0.6651	-0.851	0.1788	-0.6721
1/Competition	(0.0113)***	(0.0156)***	(0.0202)***	(0.0091)***	(0.0201)***	(0.0222)***	(0.0135)***	(0.0266)***	(0.0276)***
log **of**	-0.1176	0.0094	-0.1082	-0.0865	0.0098	-0.0767	-0.0079	0.0016	-0.0062
1/Diversity_HHI	(0.0241)***	(0.0034)**	(0.0222)***	(0.0169)***	(0.003)**	(0.0149)***	(0.0209)	-0.0044	(0.0164)
log **of**	0.5261	-0.0422	0.4838	0.539	-0.0611	0.4779	0.6623	-0.1392	0.5231
City_ind_E	(0.0225)***	(0.0118)***	(0.0225)***	(0.0176)***	(0.0145)***	(0.0218)***	(0.0111)***	(0.0207)***	(0.0215)***
log **of**	0.0835	-0.0067	0.0768	0.0511	-0.0058	0.0453	-0.0191	0.004	-0.0151
City_E	(0.0187)***	(0.0022)**	(0.0175)***	(0.015)***	(0.0022)*	(0.0133)***	(0.0133)	(0.0027)	(0.0107)
log **of**	-0.032	0.0025	-0.0294	0.0117	-0.0013	0.0104	0.0376	-0.0079	0.0297
CsumYr_pat +1	(0.0051)***	(0.0008)**	(0.0047)***	(0.0066)	(0.0008)	(0.0058)	(0.01)***	(0.0025)**	(0.0079)***

Table 7.44: Spatial Durbin Error Model Results Part1

(SDEM)Spatial panel fixed effects GM model part 1										
spgm(formula =log(LQ_size) log(LQ_general) + log(1/Competition_HHI) + log(1/Diversity_HHI) + log(City_ind_E) + log(City_E) + log(CsumYr_pat + 1) + log(LQ_general_SL) + log(1/Competition_HHI_SL) + log(1/Diversity_HHI_SL) + log(City_ind_E_SL + 1) + log(City_E_SL + 1) +log(CsumYr_pat_SL + 1), listw = Knn_neighbor_weights,model = "within", lag = T, spatial.error = T, moments = "fullweights", lag.instruments = T, method = "w2sls")										
	Sector 13		Sector 14		Sector 15		Sector 16		Sector 17	
log(**LQ_general)**	0.128	(0.018)***	0.148	(0.019)***	0.184	(0.019)***	-0.137	(0.02)***	0.353	(0.022)***
log(**1/Competition_HHI)**	-0.468	(0.006)***	-0.607	(0.007)***	-0.64	(0.009)***	-1.031	(0.02)***	-0.517	(0.009)***
log(**1/Diversity_HHI)**	-0.104	(0.013)***	-0.098	(0.015)***	-0.138	(0.015)***	-0.001	(0.021)	-0.169	(0.018)***
log(**City_ind_E)**	0.466	(0.018)***	0.512	(0.018)***	0.51	(0.019)***	1.038	(0.02)***	0.29	(0.021)***
log(**City_E)**	0.053	(0.017)**	0.035	(0.014)*	0.077	(0.017)***	-0.071	(0.015)***	0.12	(0.017)***
log(**CsumYr_pat + 1)**	-0.017	(0.004)***	-0.032	(0.005)***	-0.024	(0.007)***	-0.057	(0.006)***	-0.005	(0.007)
log(**LQ_genera_SL)**	0.024	(0.021)	-0.033	(0.03)	0.143	(0.023)***	-0.101	(0.025)***	0.127	(0.043)**
log(**1/Competition_HHI_SL)**	-0.039	(0.018)*	0.008	(0.026)	-0.006	(0.03)	0.147	(0.075)*	0.129	(0.03)***
log(**1/Diversity_HHI_SL)**	-0.19	(0.037)***	-0.124	(0.038)**	-0.08	(0.037)*	0.158	(0.055)**	-0.145	(0.047)**
log(**City_ind_E_SL+1)**	-0.041	(0.021)*	0.023	(0.028)	-0.019	(0.027)	0.112	(0.025)***	-0.128	(0.034)***
log(**City_E_SL+1)**	0.057	(0.026)*	-0.045	(0.034)	0.054	(0.029)	-0.101	(0.03)***	0.042	(0.044)
log(**CsumYr_pat + 1)**	0.02	(0.007)**	-0.003	(0.007)	0.048	(0.012)***	-0.029	(0.001)**	0.114	(0.013)***
	Sector 18		Sector 19		Sector 20		Sector 21		Sector 22	
log(**LQ_general)**	0.161	(0.02)***	0.14	(0.017)***	0.199	(0.019)***	0.241	(0.014)***	0.28	(0.02)***
log(**1/Competition_HHI)**	-0.681	(0.008)***	-0.723	(0.01)***	-0.651	(0.008)***	-0.774	(0.008)***	-0.59	(0.008)***
log(**1/Diversity_HHI)**	-0.07	(0.018)***	-0.031	(0.016)	-0.1	(0.017)***	-0.076	(0.016)***	-0.203	(0.016)***
log(**City_ind_E)**	0.561	(0.02)***	0.639	(0.017)***	0.517	(0.019)***	0.528	(0.014)***	0.455	(0.019)***
log(**City_E)**	0.022	(0.015)	-0.034	(0.014)*	0.032	(0.015)*	0.046	(0.012)***	0.058	(0.015)***
log(**CsumYr_pat + 1)**	-0.011	(0.008)	0.01	(0.009)	0.015	(0.007)*	-0.002	(0.008)	-0.018	(0.006)**
log(**LQ_genera_SL)**	-0.345	(0.04)***	-0.025	(0.015)	-0.042	(0.027)	0.028	(0.018)	0.412	(0.083)***
log(**1/Competition_HHI_SL)**	0.311	(0.057)***	-0.106	(0.04)**	0.269	(0.035)***	-0.007	(0.031)	0.428	(0.067)***
log(**1/Diversity_HHI_SL)**	0.139	(0.067)*	-0.021	(0.034)	-0.228	(0.053)***	-0.085	(0.042)*	-0.125	(0.11)
log(**City_ind_E_SL+1)**	0.077	(0.028)**	0.01	(0.015)	-0.02	(0.022)	-0.025	(0.015)	-0.309	(0.073)***
log(**City_E_SL+1)**	-0.243	(0.043)***	-0.026	(0.025)	-0.156	(0.041)***	0.026	(0.031)	0.159	(0.079)*
log(**CsumYr_pat + 1)**	-0.041	(0.017)*	0.02	(0.011)	0.07	(0.017)***	0.004	(0.013)	0.036	(0.019)
	Sector 23		Sector 24		Sector 25		Sector 26		Sector 27	
log(**LQ_general)**	0.222	(0.015)***	0.071	(0.022)**	0.323	(0.023)***	0.303	(0.019)***	0.131	(0.019)***
log(**1/Competition_HHI)**	-0.744	(0.008)***	-0.727	(0.011)***	-0.681	(0.011)***	-0.429	(0.007)***	-0.553	(0.008)***
log(**1/Diversity_HHI)**	-0.091	(0.014)***	-0.07	(0.023)**	-0.135	(0.024)***	-0.225	(0.014)***	-0.085	(0.014)***
log(**City_ind_E)**	0.479	(0.015)***	0.69	(0.022)***	0.507	(0.023)***	0.352	(0.019)***	0.539	(0.018)***
log(**City_E)**	0.072	(0.011)***	-0.035	(0.016)*	0.083	(0.018)***	0.076	(0.016)***	0.025	(0.016)
log(**CsumYr_pat + 1)**	0.003	(0.007)	0.007	(0.007)	0.031	(0.009)***	-0.025	(0.004)***	-0.013	(0.004)**
log(**LQ_genera_SL)**	-0.037	(0.03)	-0.082	(0.029)**	0.162	(0.027)***	0.167	(0.035)***	-0.042	(0.025)
log(**1/Competition_HHI_SL)**	-0.149	(0.027)***	0.073	(0.039)	0.19	(0.047)***	0.123	(0.028)***	-0.036	(0.031)
log(**1/Diversity_HHI_SL)**	0.008	(0.035)	0.022	(0.061)	-0.199	(0.068)**	-0.3	(0.055)***	-0.221	(0.035)***
log(**City_ind_E_SL+1)**	0.055	(0.027)*	0.111	(0.025)***	-0.142	(0.026)***	-0.136	(0.034)***	-0.04	(0.027)
log(**City_E_SL+1)**	-0.116	(0.033)***	-0.121	(0.039)**	0.065	(0.035)	0.08	(0.041)	0.012	(0.03)
log(**CsumYr_pat + 1)**	-0.014	(0.012)	-0.01	(0.013)	0.013	(0.017)	0.06	(0.008)***	0.033	(0.006)***

Table 7.45: Spatial Durbin Error Model Results Part 2

(SDEM)Spatial panel fixed effects GM model part 2										
	Sector 28		Sector 29		Sector 30		Sector 31		Sector 32	
log(**LQ_general)**	0.287	(0.027)***	0.028	(0.021)***	0.206	(0.02)***	0.199	(0.018)***	0.234	(0.024)***
log(**1/Competition_HHI)**	-0.718	(0.014)***	-0.65	(0.001)***	-0.581	(0.008)***	-0.431	(0.007)***	-0.61	(0.009)***
log(**1/Diversity_HHI)**	-0.11	(0.03)***	-0.17	(0.02)***	-0.12	(0.016)***	-0.182	(0.012)***	-0.142	(0.02)***
log(**City_ind_E)**	0.515	(0.026)***	0.494	(0.021)***	0.483	(0.02)***	0.347	(0.017)***	0.491	(0.024)***
log(**City_E)**	0.056	(0.021)**	0.059	(0.017)***	0.026	(0.016)	0.097	(0.014)***	0.065	(0.02)**
log(**CsumYr_pat + 1)**	-0.016	(0.009)	-0.01	(0.006)	-0.024	(0.005)***	-0.014	(0.004)**	-0.043	(0.007)***
log(**LQ_genera_SL)**	0.107	(0.04)**	0.049	(0.039)	-0.28	(0.047)***	0.196	(0.028)***	0.128	(0.05)*
log(**1/Competition_HHI_SL)**	0.406	(0.083)***	0.56	(0.051)***	0.071	(0.048)	0.103	(0.022)***	0.667	(0.08)***
log(**1/Diversity_HHI_SL)**	-0.264	(0.091)**	-0.126	(0.051)*	-0.321	(0.109)**	-0.16	(0.039)***	-0.304	(0.124)*
log(**City_ind_E_SL+1)**	-0.208	(0.032)***	-0.078	(0.044)*	-0.046	(0.036)	-0.203	(0.029)***	-0.108	(0.052)*
log(**City_E_SL+1)**	0.244	(0.05)***	-0.164	(0.048)***	-0.034	(0.052)	0.051	(0.033)	0.117	(0.054)*
log(**CsumYr_pat + 1)**	0.148	(0.017)***	0.031	(0.011)**	0.03	(0.014)*	0.044	(0.008)***	0.005	(0.016)
	Sector 33		Sector 34		Sector 35		Sector 36		Sector 37	
log(**LQ_general)**	0.133	(0.017)***	0.228	(0.02)***	0.25	(0.022)***	0.258	(0.02)***	0.151	(0.022)***
log(**1/Competition_HHI)**	-0.624	(0.009)***	-0.571	(0.008)***	-0.515	(0.009)***	-0.569	(0.009)***	-0.559	(0.009)***
log(**1/Diversity_HHI)**	-0.128	(0.02)***	-0.171	(0.017)***	-0.181	(0.018)***	-0.142	(0.017)***	-0.122	(0.019)***
log(**City_ind_E)**	0.593	(0.016)***	0.463	(0.019)***	0.39	(0.021)***	0.436	(0.019)***	0.533	(0.021)***
log(**City_E)**	0.047	(0.015)**	0.046	(0.015)**	0.069	(0.018)***	0.12	(0.016)***	0.029	(0.017)
log(**CsumYr_pat + 1)**	-0.01	(0.006)	-0.011	(0.005)*	-0.052	(0.005)***	-0.051	(0.005)***	-0.082	(0.005)***
log(**LQ_genera_SL)**	-0.003	(0.039)	-0.102	(0.037)**	-0.192	(0.061)**	0.219	(0.053)***	0.088	(0.046)
log(**1/Competition_HHI_SL)**	0.334	(0.05)***	0.051	(0.036)	0.126	(0.036)***	-0.069	(0.037)	-0.195	(0.037)***
log(**1/Diversity_HHI_SL)**	-0.282	(0.107)**	-0.197	(0.065)**	-0.361	(0.078)***	-0.144	(0.056)*	-0.158	(0.052)**
log(**City_ind_E_SL+1)**	-0.003	(0.031)	0.096	(0.037)**	0.021	(0.053)	-0.147	(0.036)***	0.032	(0.034)
log(**City_E_SL+1)**	-0.087	(0.052)	-0.245	(0.051)***	-0.051	(0.071)	0.007	(0.059)	0.055	(0.04)
log(**CsumYr_pat + 1)**	0.076	(0.015)***	0.022	(0.012)	0.095	(0.009)***	0.145	(0.01)***	0.099	(0.007)***
	Sector 39		Sector 40		Sector 41		Sector 42		Sector 43	
log(**LQ_general)**	0.192	(0.02)***	0.297	(0.022)***	0.238	(0.023)***	0.243	(0.019)***	0.156	(0.01)***
log(**1/Competition_HHI)**	-0.538	(0.008)***	-0.616	(0.01)***	-0.726	(0.011)***	-0.747	(0.008)***	-0.855	(0.012)***
log(**1/Diversity_HHI)**	-0.135	(0.017)***	-0.174	(0.025)***	-0.134	(0.024)***	-0.094	(0.017)***	-0.02	(0.022)***
log(**City_ind_E)**	0.499	(0.02)***	0.452	(0.021)***	0.525	(0.022)***	0.533	(0.02)***	0.669	(0.011)***
log(**City_E)**	0.014	(0.015)	0.097	(0.018)***	0.07	(0.019)***	0.046	(0.014)**	-0.009	(0.013)**
log(**CsumYr_pat + 1)**	-0.035	(0.004)***	-0.044	(0.005)***	-0.041	(0.006)***	0.007	(0.006)	0.034	(0.01)
log(**LQ_genera_SL)**	-0.32	(0.061)***	0.075	(0.042)	0.105	(0.033)**	-0.134	(0.039)***	-0.055	(0.022)***
log(**1/Competition_HHI_SL)**	0.184	(0.047)***	-0.122	(0.091)	0.105	(0.035)**	0.212	(0.057)***	-0.162	(0.062)***
log(**1/Diversity_HHI_SL)**	-0.073	(0.109)	-0.349	(0.115)**	-0.149	(0.059)*	-0.122	(0.069)	-0.251	(0.094)
log(**City_ind_E_SL+1)**	0.363	(0.058)***	-0.135	(0.043)**	-0.093	(0.03)**	0.05	(0.028)	-0.026	(0.027)
log(**City_E_SL+1)**	-0.524	(0.085)***	0.07	(0.073)	0.05	(0.039)	-0.124	(0.047)**	0.005	(0.048)**
log(**CsumYr_pat + 1)**	-0.016	(0.008)	0.057	(0.016)***	0.02	(0.007)**	0.023	(0.014)	-0.071	(0.026)

Chapter 8

Appendix to Chapter 4

8.1 Top 10 ranked with betweenness centrality and HITS for co-patenting and citation networks from 1998 to 2013

Table 8.1: Top 10 ranked with betweenness centrality in co-patenting network

Rank	1998	1999	2000	2001	2002	2003	2004	2005
1	C01	C01	C01	C01	C01	C01	C01	U01
2	R01	R01	U03	U01	U01	U01	U01	C01
3	U12	U12	R01	R01	R01	R01	U03	U03
4	C03	C04	U01	U03	C03	C03	R01	U04
5	C04	C03	C03	C03	U03	U03	U04	C03
6	U03	U01	U12	U02	U12	U12	C03	U06
7	U01	U03	C04	U12	U02	U04	U06	U11
8	R02	R02	U13	U09	C04	U02	U11	R01
9	U13	U13	U02	C04	U09	U11	U02	U12
10	U02	U02	U09	U10	U04	U08	U12	U08
Rank	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
1	U01	U01	U01	U01	U01	U01	U01	U01
2	C01	C01	C01	C01	C01	U03	U03	C05
3	U03	U03	U03	U03	U03	C01	U04	U03
4	C03	C03	U04	U04	U04	U04	C01	U04
5	U04	U04	C03	C03	C03	U05	C05	C01
6	U06	U06	U06	U06	U05	C02	U05	U05
7	U12	U11	C06	U05	U06	U07	U02	U02
8	U11	C06	U11	U11	U11	U06	U07	U07
9	U02	U12	U12	C06	U02	C03	U06	C02
10	R01	U02	U02	U12	U12	U02	C01	U08

Table 8.2: Name list for table 5.1

University	Chinese Names	English Names
U01	清华大学	Tsinghua University
U02	北京大学	Peking University
U03	浙江大学	Zhejiang University
U04	上海交通大学	Shanghai Jiao Tong University
U05	华南理工大学	South China University of Technology
U06	华东理工大学	East China University of Science and Technology
U07	东华大学	Donghua University
U08	四川大学	Sichuan University
U09	中南大学	Central South University
U10	大连理工大学	Dalian University of Technology
U11	复旦大学	Fudan University
U12	北京科技大学	University of Science and Technology Beijing
U13	天津大学	Tianjin University
Company	**Chinese Names**	**English Names**
C01	中国石油化工股份有限公司	Sinopec
C02	中国石油天然气股份有限公司	PetroChina
C03	上海宝山钢铁总厂	Baosteel Group
C04	鞍山钢铁公司	Ansteel Group
C05	国家电网公司	State Grid Corporation of China
C06	深圳市华为技术有限公司	Huawei Technology
Research Institute	**Chinese Names**	**English Names**
R01	冶金工业部钢铁研究总院	Central Iron & Steel Research Institute
R02	中国科学院过程工程研究所	Institute of Process Engineering,Chinese Academy of Sciences

Table 8.3: Authority's Rank of Each Year

Authority's Rank of Each Year

1998	1999	2000	2001	2002	2003	2004	2005
US01	US01	US05	US05	US05	US12	JP04	KR01
US02	US05	US01	US01	CN02	CN02	KR01	JP04
US03	US02	FR01	CN02	US01	US14	DE02	US12
US04	US03	CN02	FR01	FR01	JP08	JP01	JP05
US05	JP03	JP03	JP03	US09	JP04	US12	DE02
US06	US06	US07	US07	JP03	CN06	SE01	JP01
JP02	FR01	DE01	US10	US07	JP01	JP05	US14
FR01	US04	US02	JP02	US10	US13	US13	JP06
CN01	US15	US08	US17	JP02	JP05	US14	US18
JP03	US16	NL01	US08	US11	KR01	JP06	KR02
2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
KR01	KR01	KR01	KR01	CN03	TW01	TW01	TW01
JP04	JP04	JP04	JP04	KR01	CN04	CN04	CN04
JP05	JP05	JP05	CN03	CN04	CN03	CN03	CN03
JP01	US14	JP01	CN04	TW01	KR01	KR01	CN05
US12	NL02	US14	JP05	JP04	CN05	CN05	KR01
DE02	JP01	JP06	TW01	CN05	JP04	JP04	JP04
JP06	US12	JP07	JP01	JP05	JP06	CN06	CN06
JP07	JP07	CN03	JP06	JP06	JP05	JP06	JP06
KR02	DE02	NL02	CN05	JP01	JP01	JP01	JP07
US14	JP06	US12	JP07	US14	CN06	US12	US12

Table 8.4: Hub's Rank of Each Year

1998	1999	2000	2001	2002	2003	2004	2005
CN02	CN02	CN02	CN02	CN02	CN03	CN03	CN03
CN07	CN01	CN01	CN01	CN01	CN02	TW01	TW01
CN08	CN08	CN10	CN10	CN10	CN01	CN05	CN04
CN09	CN07	CN08	CN08	CN03	CN10	CN04	CN05
CN01	CN10	CN07	CN07	CN07	CN05	TW02	CN21
CN10	CN09	CN09	CN16	CN08	CN17	CN20	KR01
CN11	CN11	CN11	CN09	CN16	CN06	FR02	TW02
CN12	CN14	CN15	CN15	CN09	CN18	CN19	CN20
CN13	CN12	CN12	CN11	CN15	CN19	KR01	FR02
JP09	CN15	CN14	CN12	CN11	TW01	CN18	CN06
2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
CN03	CN03	CN03	CN03	CN03	CN03	CN03	CN03
TW01	TW01	TW01	CN05	CN05	CN05	CN05	CN05
CN04	CN04	CN04	TW01	TW01	TW01	CN25	CN25
CN05	CN05	CN05	CN04	CN04	CN25	CN27	CN27
CN06	CN06	CN06	CN25	CN25	CN04	TW01	TW01
CN21	CN22	CN25	CN06	CN21	CN21	CN28	CN28
CN22	CN21	CN22	CN21	CN06	CN26	CN04	CN04
CN23	CN23	CN21	CN26	CN26	CN06	CN29	CN29
KR01	TW03	CN23	CN22	KR01	CN27	CN21	CN21
TW02	CN24	KR01	KR01	CN22	KR01	CN26	CN26

Table 8.5: Non-Chinese organizations list in table 5.3 and 5.4

Code	English Names	Code	English Names
US01	Honeywell UOP	TW03	Innolux Corporation
US02	MOBIL OIL CORPORATION	SE01	Telefonaktiebolaget LM Ericsson AB
US03	DUPONT	NL01	MONTELL TECHNOLOGY COMPANY
US04	W. R. GRACE and COMPANY	NL02	KININKLIJKE PHILIPS N.V.
US05	PHILLIPS PETROLEUM COMPANY	KR01	SAMSUNG ELECTRONICS COMPANY
US06	TEXACO	KR02	LG ELECTRONICS
US07	EXXON CHEMICAL	JP01	HITACHI
US08	SHELL OIL COMPANY	JP02	AIST
US09	DOW CHEMICAL COMPANY	JP03	ASAHI KASEI KOGYO
US10	UNION CARBIDE CORPORATION	JP04	PANASONIC CORPORATION
US11	EXXON RESEARCH AND ENGINEERING COMPANY	JP05	NEC CORPORATION
US12	IBM (INTERNATIONAL BUSINESS MACHINES CORPORATION)	JP06	TOSHIBA CORPORATION
US13	MOTOROLA	JP07	SONY CORPORATION
US14	MITSUBISHI ELECTRIC CORPORATION	JP08	RICOH COMPANY
US15	CHEVRON RESEARCH COMPANY	JP09	Mitsui Chemicals
US16	ASHLAND OIL	FR01	French Institute of Petroleum (IFP Énergies nouvelles)
US17	STONE and WEBSTER	FR02	Mstar France SAS
US18	LUCENT TECHNOLOGIES	DE01	BASF
TW01	Hon Hai Precision Industry Co., Ltd.	DE02	SIEMENS
TW02	MStar Semiconductor, Inc.		

Table 8.6: Chinese organizations list in table 5.3 and 5.4

Code	English Names	Chinese Names
CN01	Research institute of Sinopec	中国石油化工总公司石油化工规划院
CN02	Sinopec Group	中国石油化工集团公司
CN03	Huawei Technology Co., Ltd.	华为技术有限公司
CN04	Hong Fu Jin Precision Industry (ShenZhen) Co.,Ltd.	鸿富锦精密工业 (深圳) 有限公司
CN05	ZTE Corporation	中兴通讯股份有限公司
CN06	Tsinghua University	清华大学
CN07	Sinopec's research institute in Fushun City	中国石油化工总公司抚顺石油化工研究所
CN08	Luoyang Petrochemical Engineering Corporation Ltd	中国石化洛阳石油化工工程公司
CN09	Sinopec Shanghai Research institute	中国石油化工总公司上海石油化工研究所
CN10	Beijing Chemical Industry Research Institute	化学工业部北京化工研究院
CN11	Sinopec Qilu Petrochemical	中国石化集团齐鲁石油化工公司
CN12	Institute of metal research, Chinese academy of sciences	中国科学院金属研究所
CN13	Daqing Petroleum Administration Bureau of PetroChina	大庆石油管理局
CN14	Sinopec Jiangsu	中国石油化工股份有限公司江苏油田分公司石油工程技术研究院
CN15	Sinopec Beijing	中国石油化工股份有限公司北京燕山分公司
CN16	Guangdong Esquel Garment Co,. Ltd.	广东溢达纺织有限公司
CN17	CapitalBio Corporation	博奥生物集团有限公司
CN18	Ruizhangtech Co,. Ltd.	瑞章科技有限公司
CN19	Datang Mobile Communication Equipment Co,. Ltd.	大唐移动通信设备有限公司
CN20	Mstar sofware research Shenzhen Co,. Ltd.	晨星软件研发 (深圳) 有限公司
CN21	Lenovo (Beijing) Limited	联想 (北京) 有限公司
CN22	Mindray Medical International Limited	深圳迈瑞生物医疗电子股份有限公司
CN23	Fuzhun Precision Industry Shenzhen Co,. Ltd.	富准精密工业 (深圳) 有限公司
CN24	Innolux Display Shenzhen Co,. Ltd.	群康科技 (深圳) 有限公司
CN25	Huawei Device Co,. Ltd.	华为终端有限公司
CN26	BYD Company	比亚迪股份有限公司
CN27	Tencent Technology (Shenzhen) Co., Ltd.	腾讯科技 (深圳) 有限公司
CN28	BOE Technology Group Co., Ltd.	北京东方电子集团股份有限公司
CN29	Shenzhen China Star Optoelectronics Technology Co., Ltd.	深圳市华星光电技术有限公司

8.2 Visualization of co-patenting network

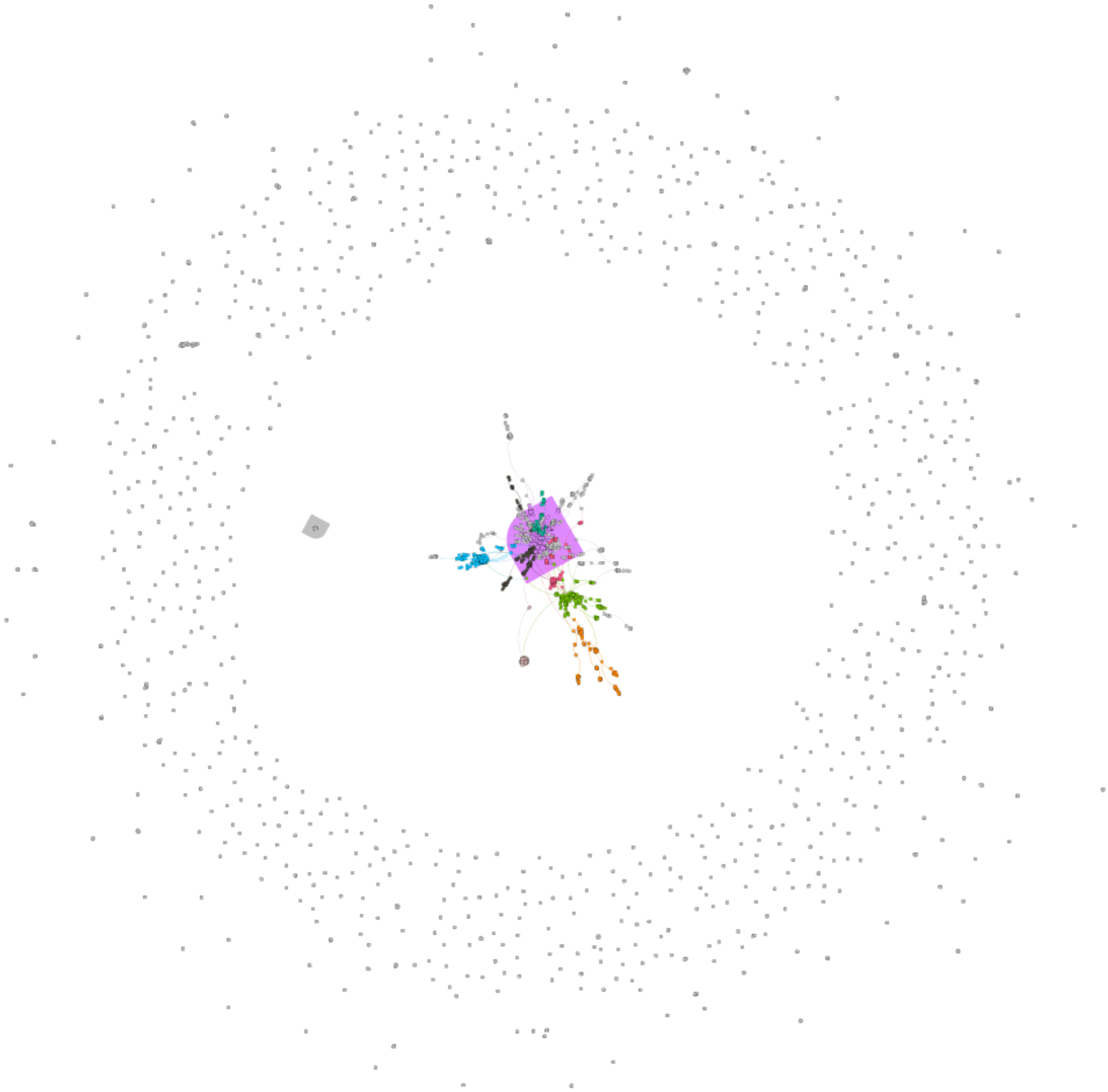


Figure 8.1: Year 2001

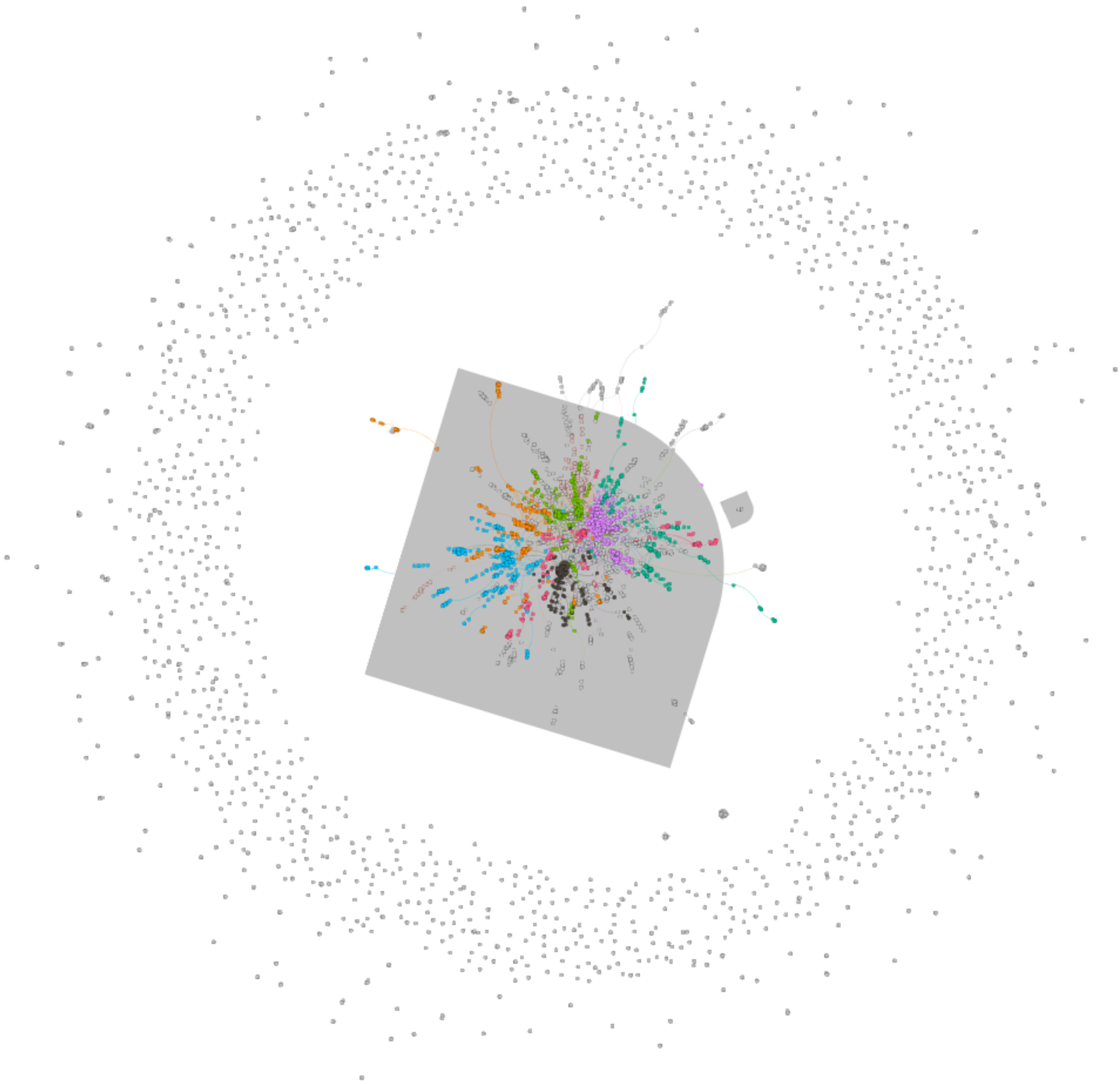


Figure 8.2: Year 2005

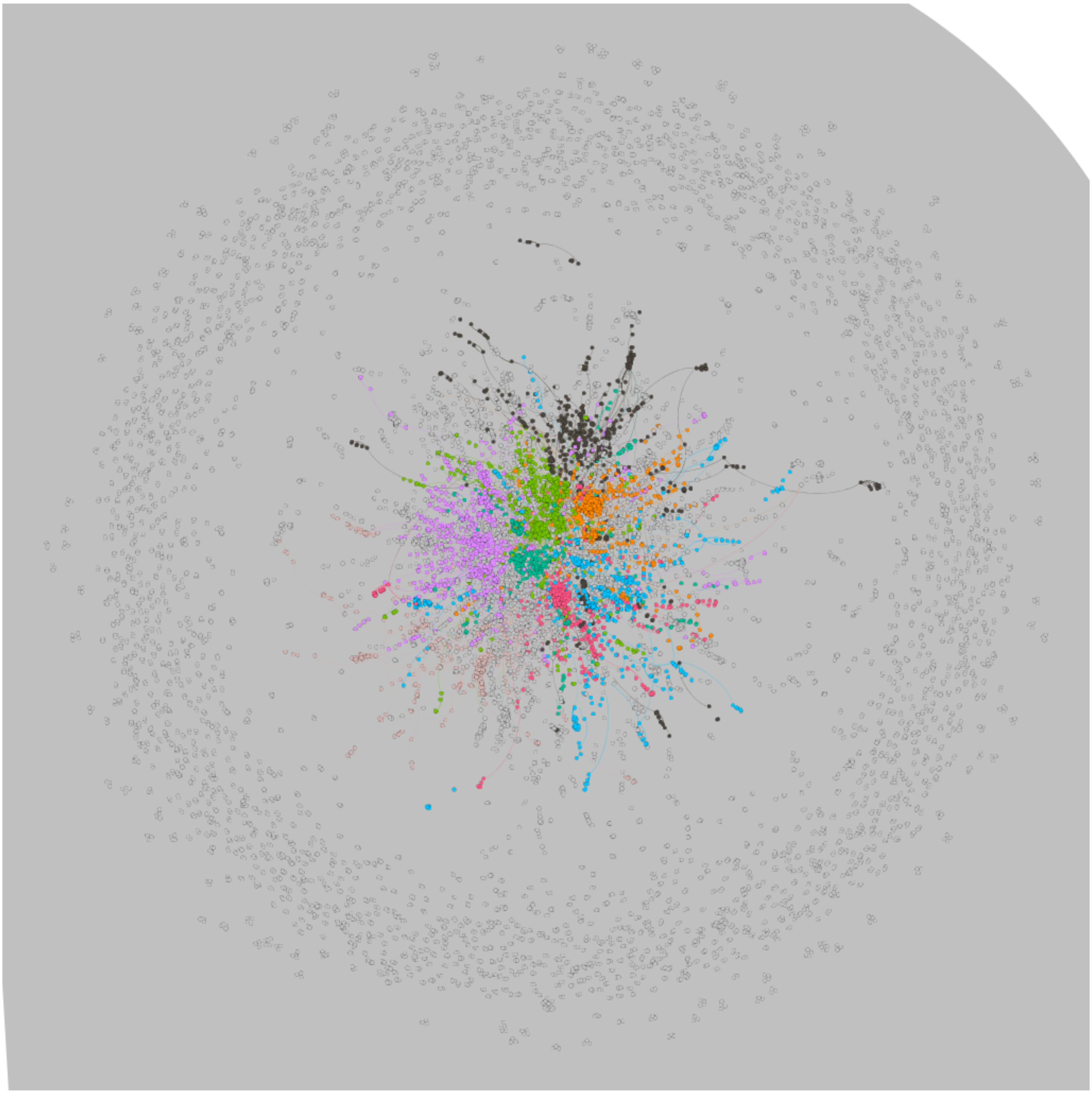


Figure 8.3: Year 2009

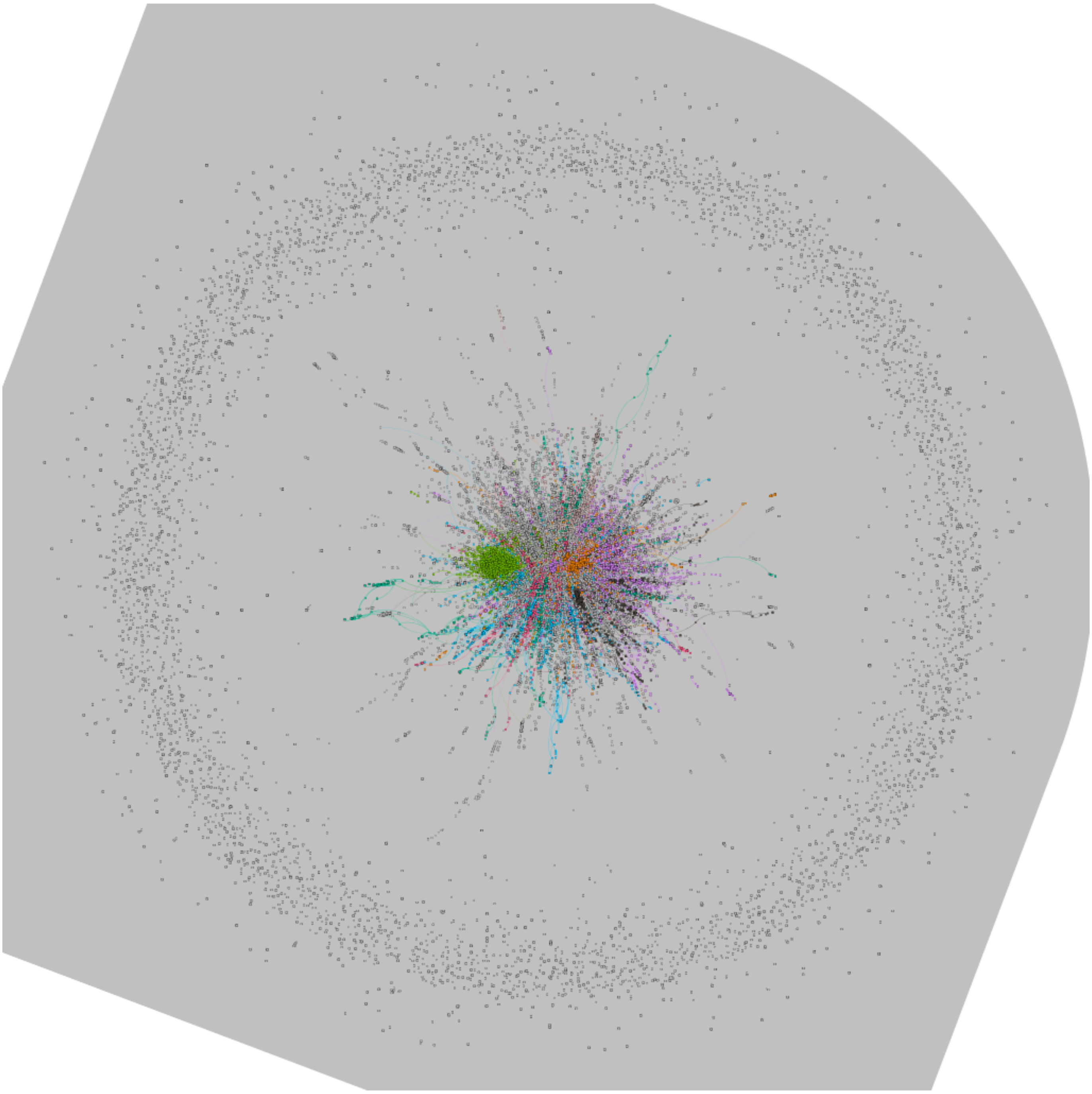


Figure 8.4: Year 2013

8.3 Visualization of citation network

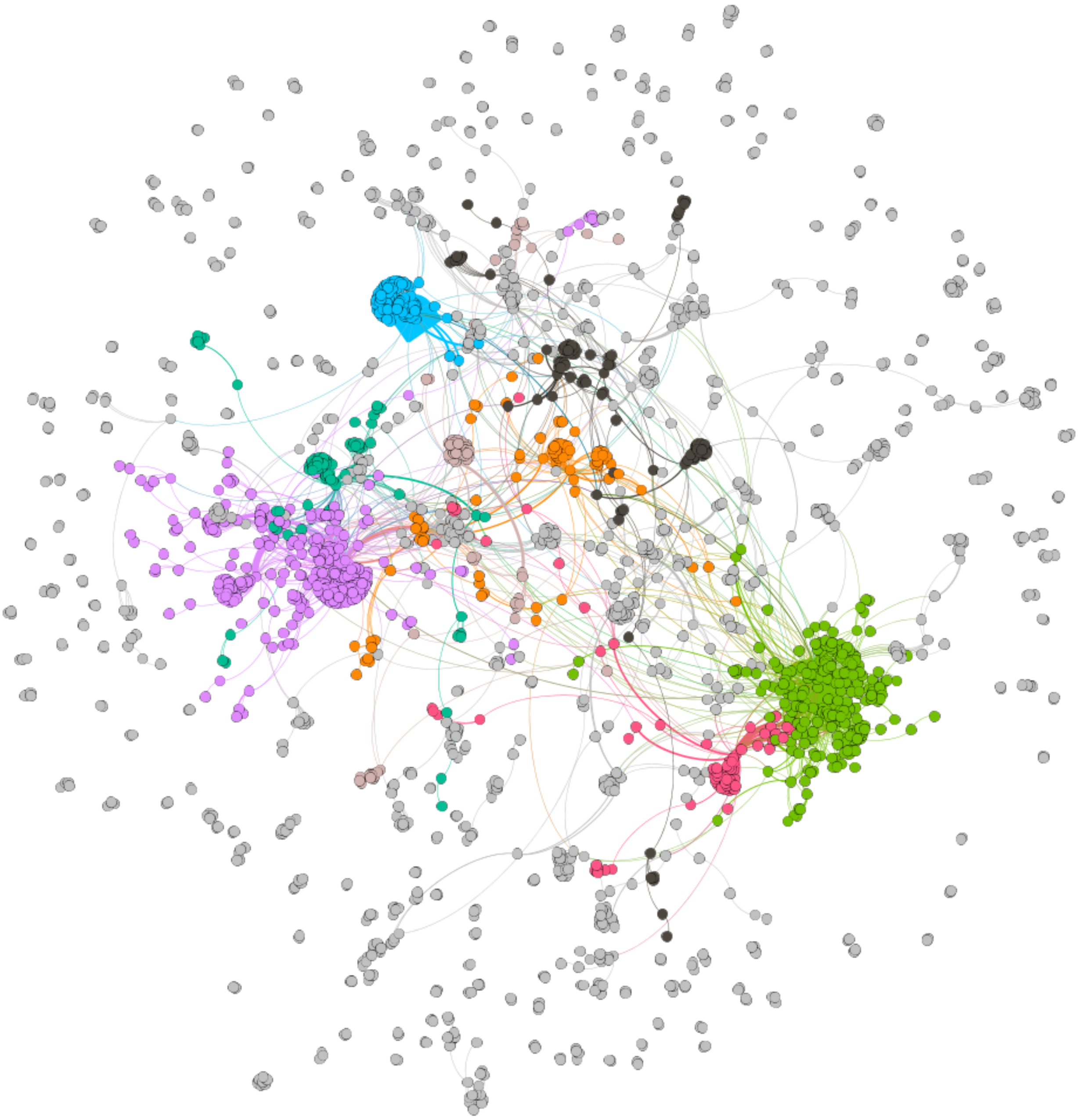


Figure 8.5: Year 2001

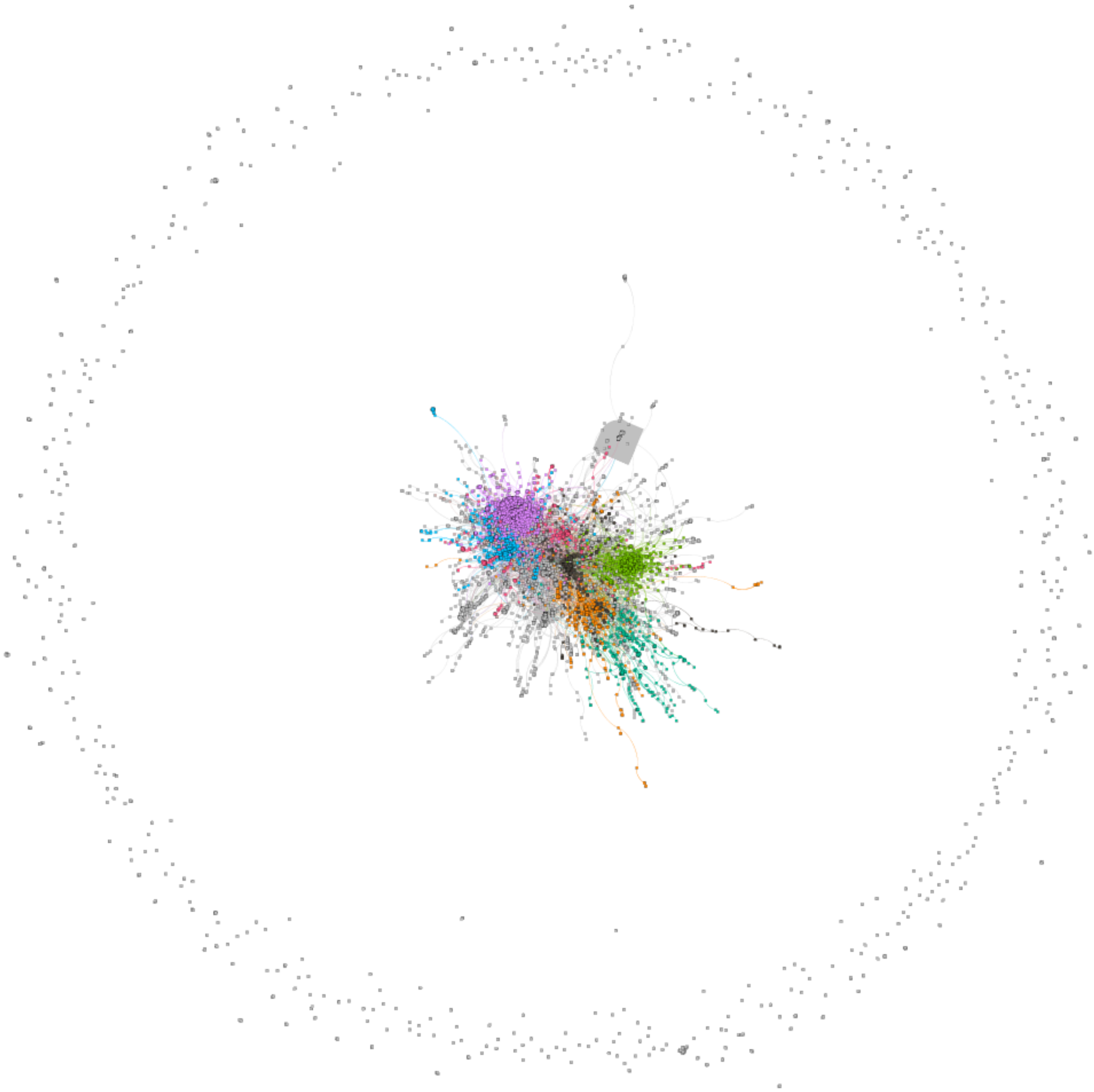


Figure 8.6: Year 2005

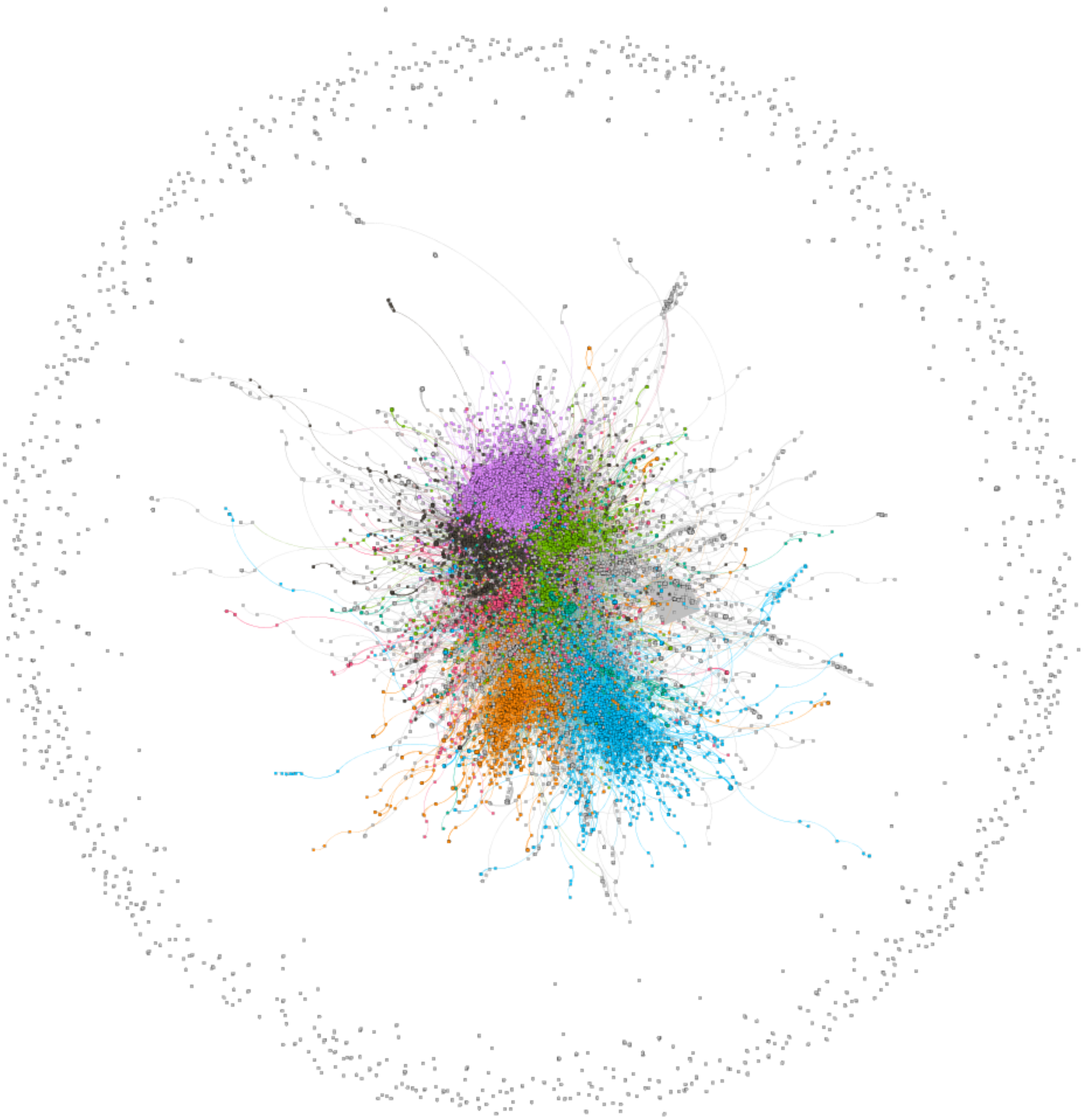


Figure 8.7: Year 2009

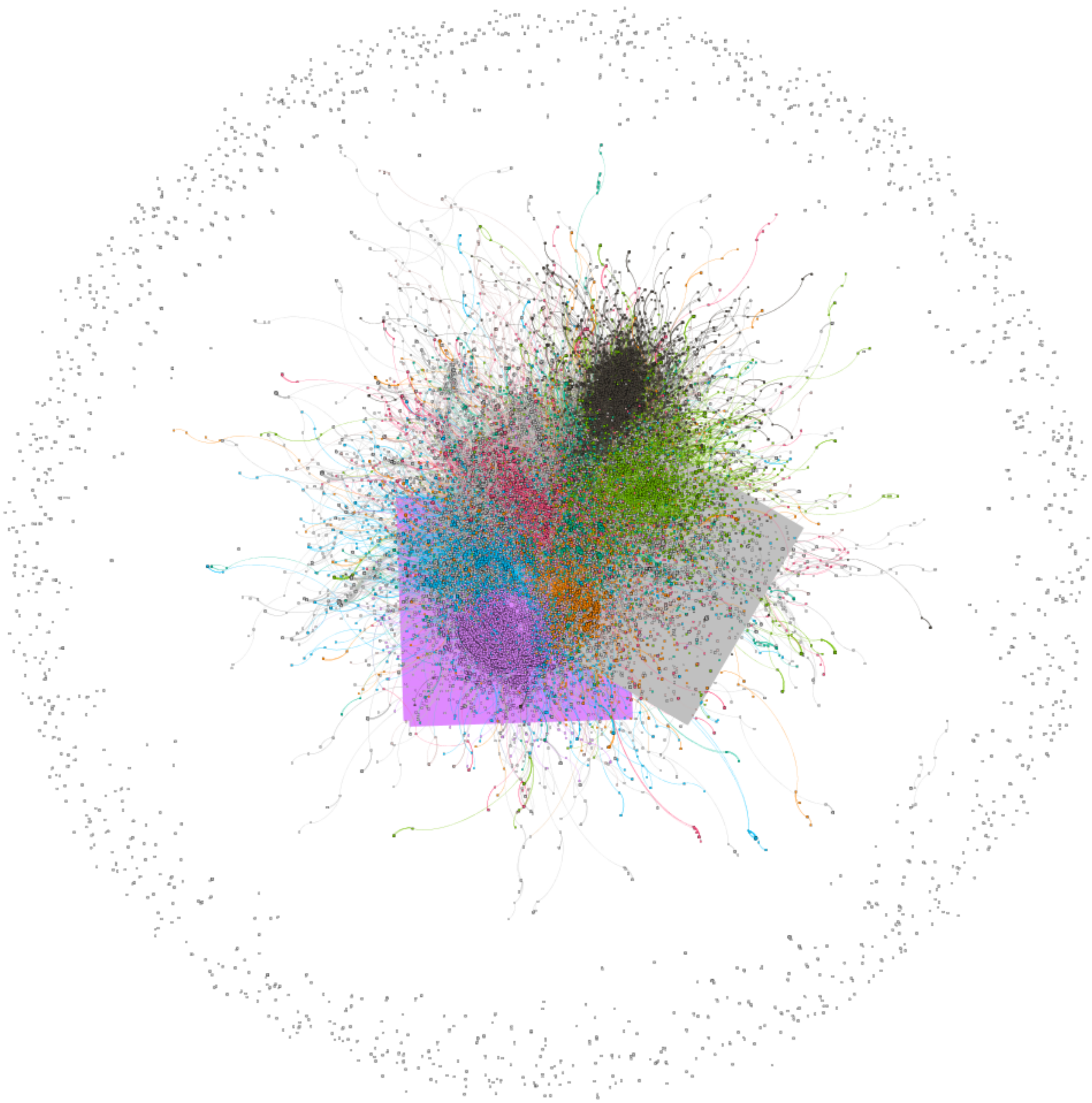


Figure 8.8: Year 2013

8.4 Yearly shift of each sector's Farrell Efficiency

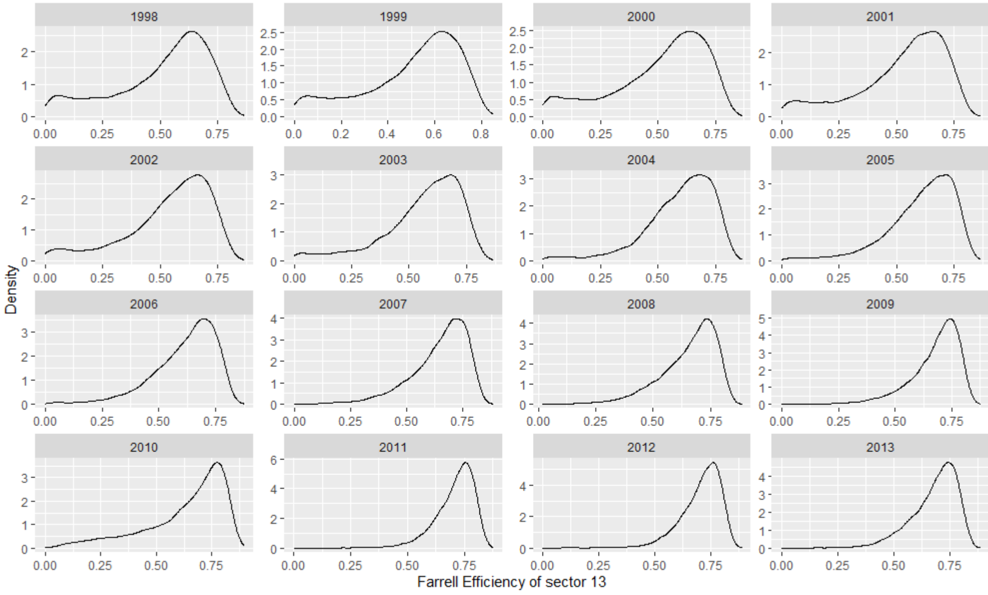


Figure 8.9: sector 13

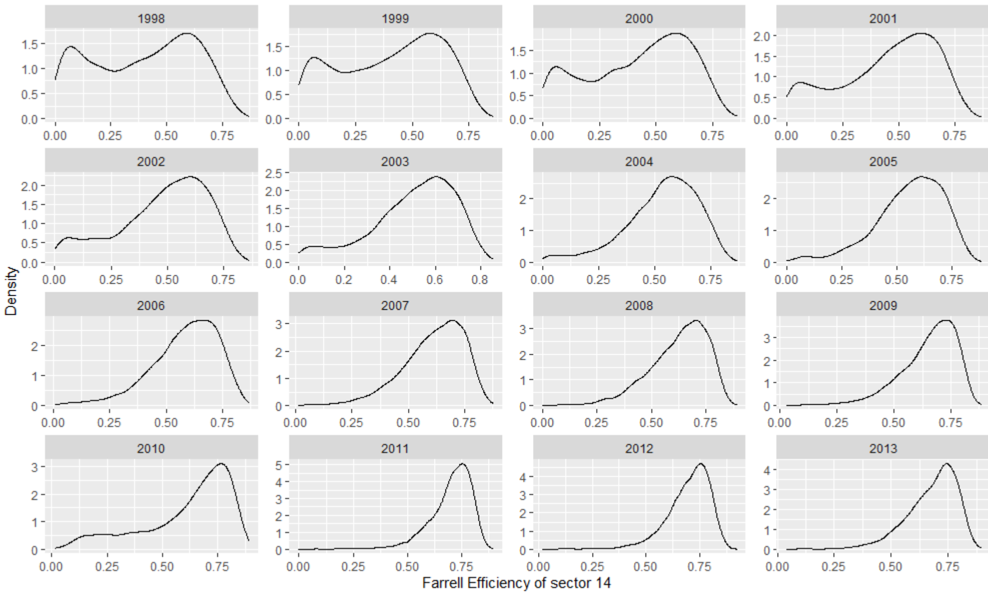


Figure 8.10: sector 14

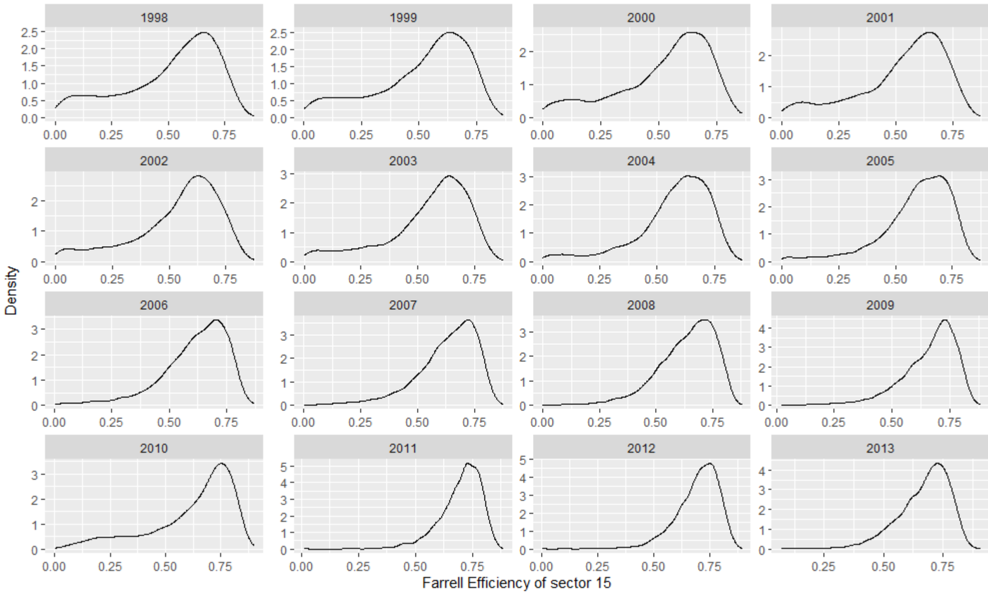


Figure 8.11: sector 15

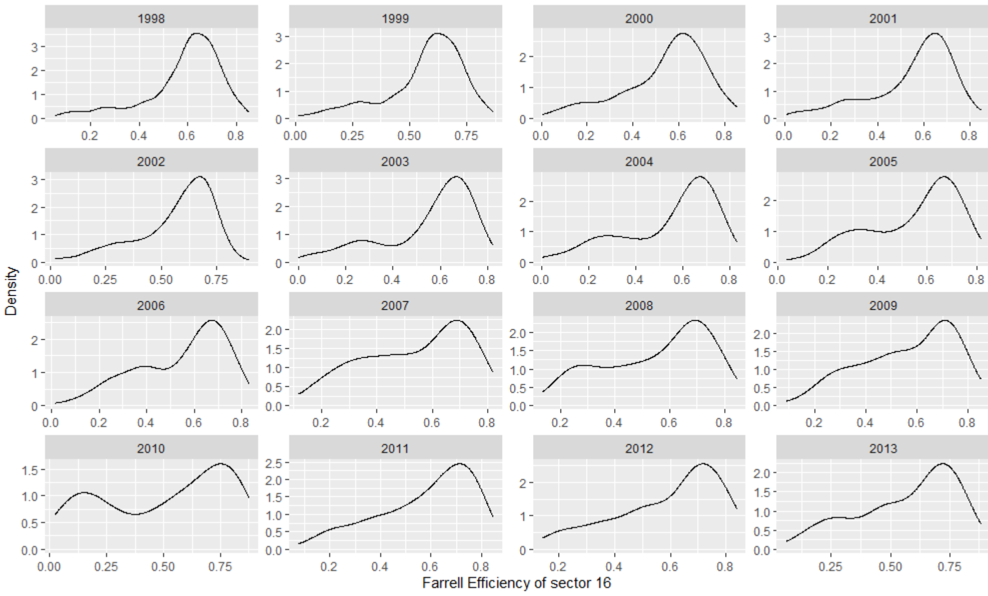


Figure 8.12: sector 16

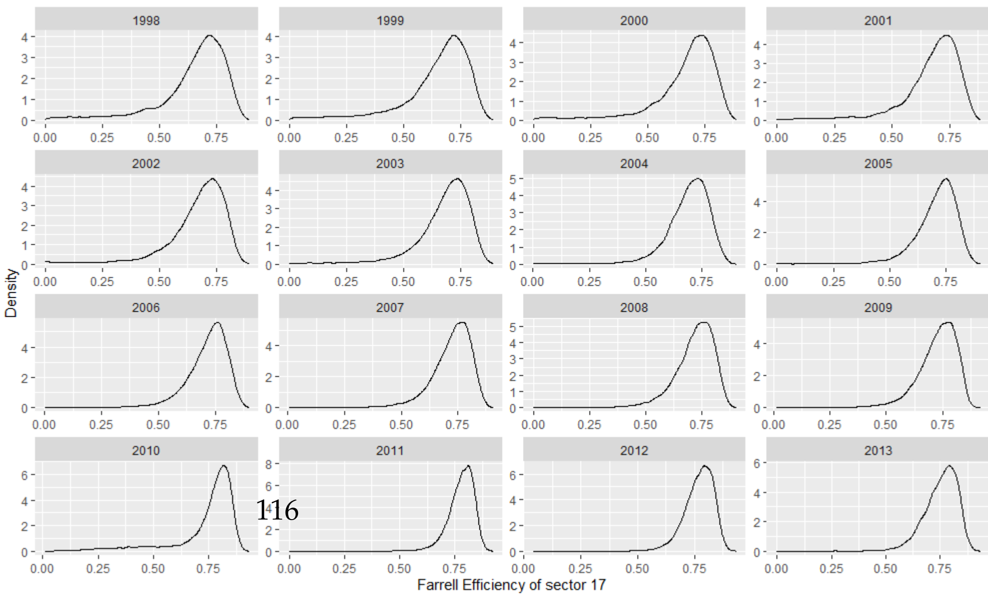


Figure 8.13: sector 17

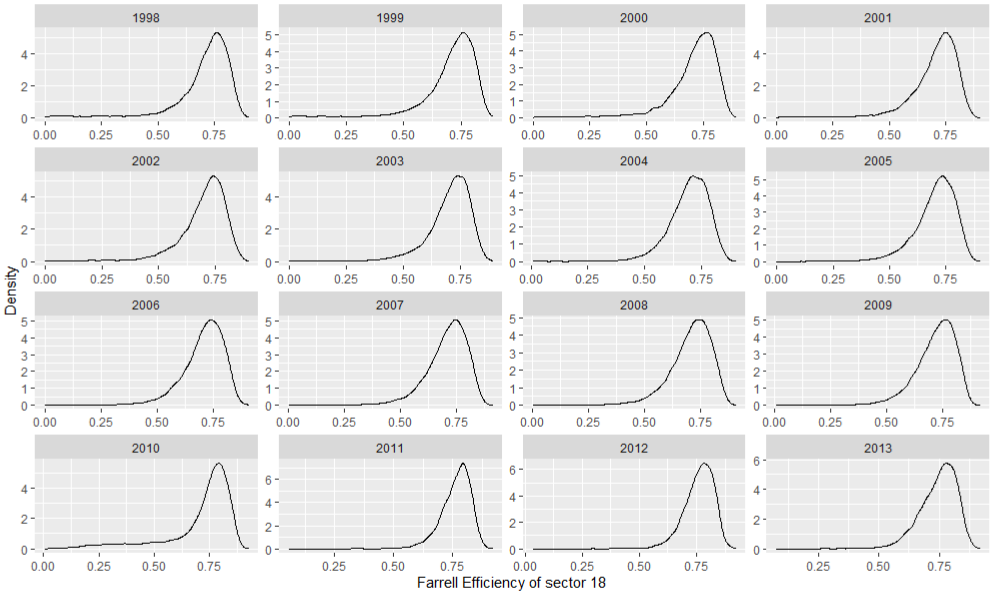


Figure 8.14: sector 18

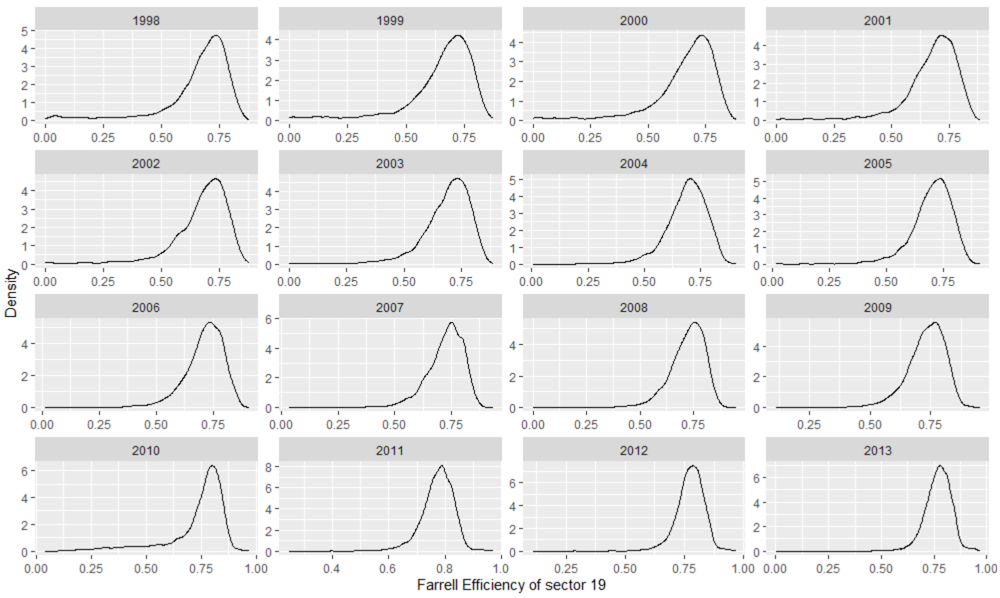


Figure 8.15: sector 19

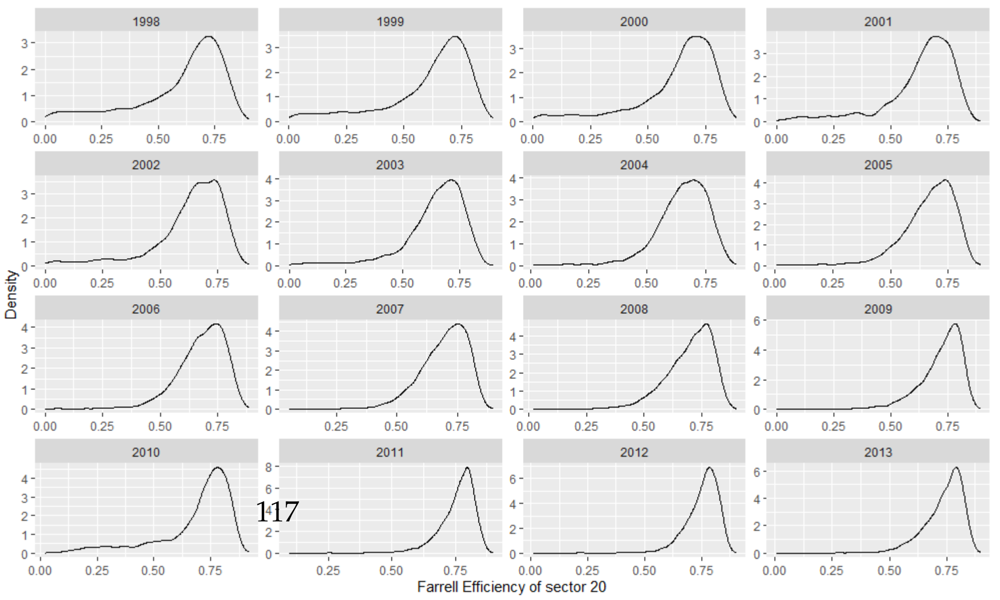


Figure 8.16: sector 20

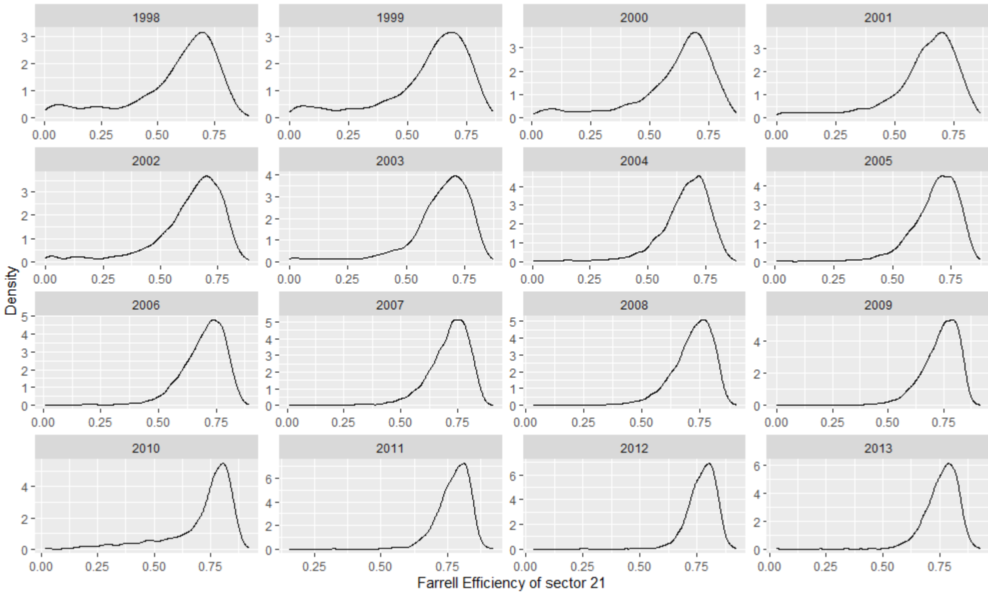


Figure 8.17: sector 21

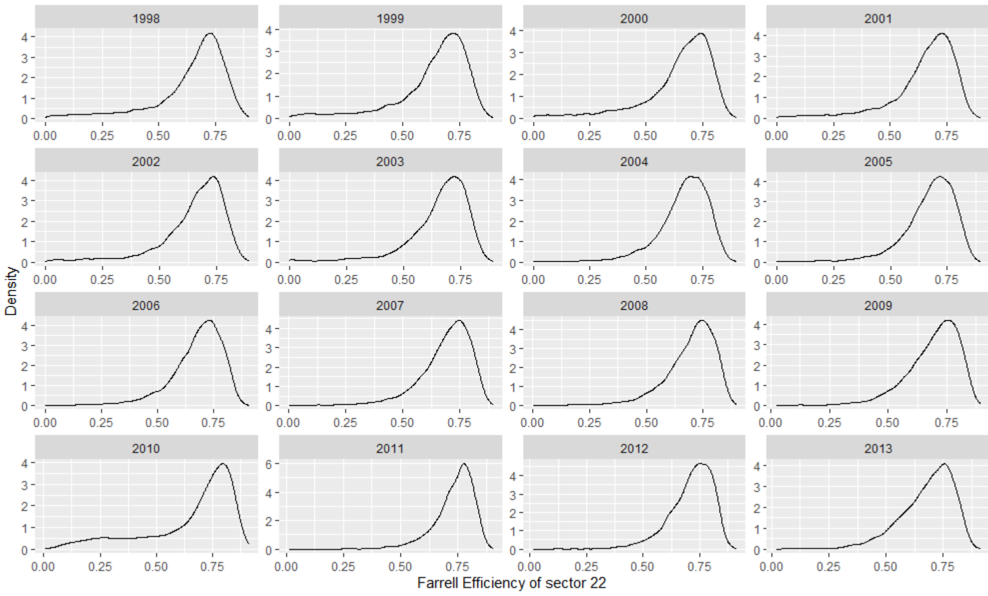


Figure 8.18: sector 22

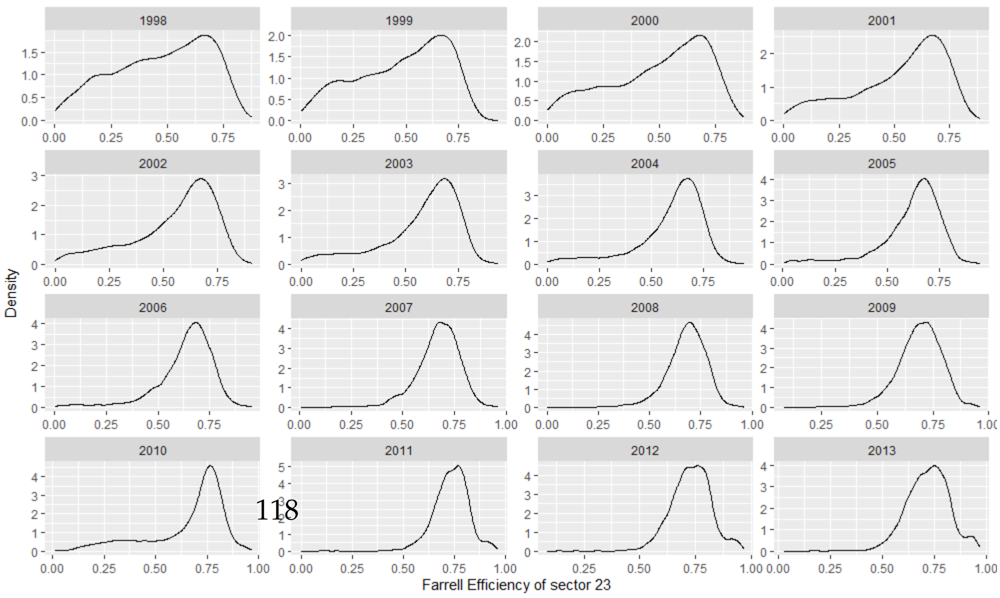


Figure 8.19: sector 23

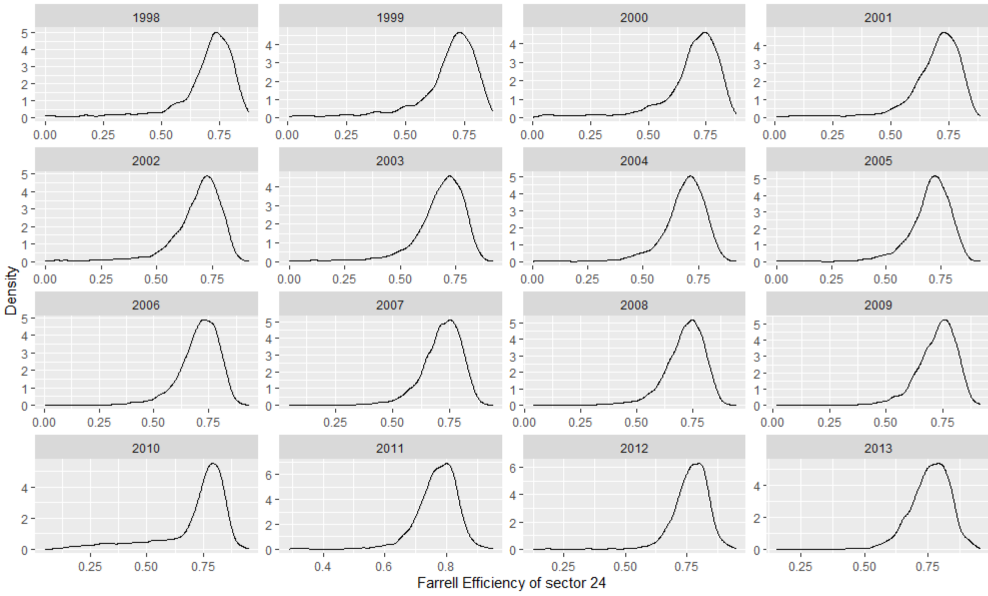


Figure 8.20: sector 24

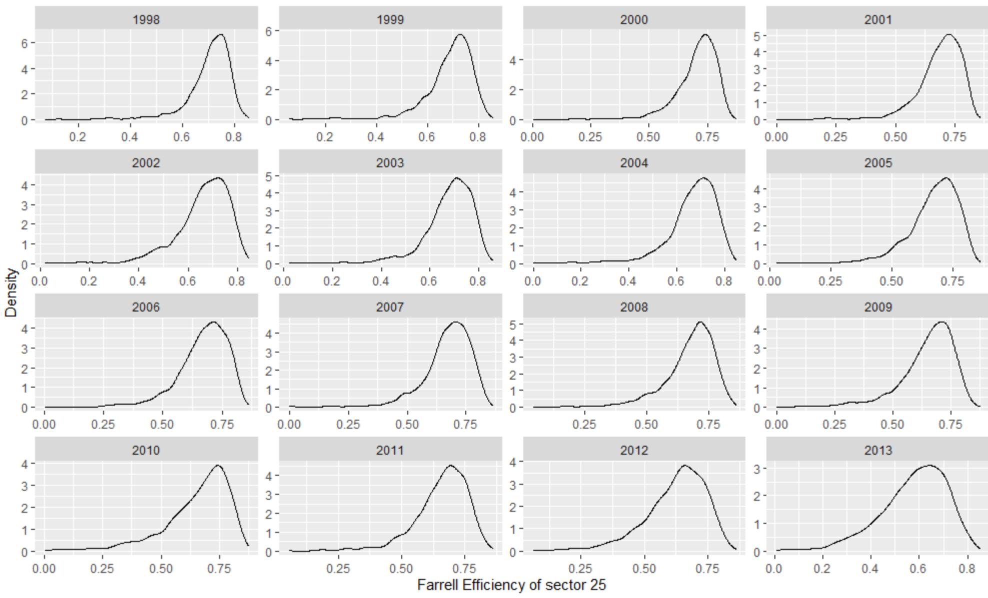


Figure 8.21: sector 25

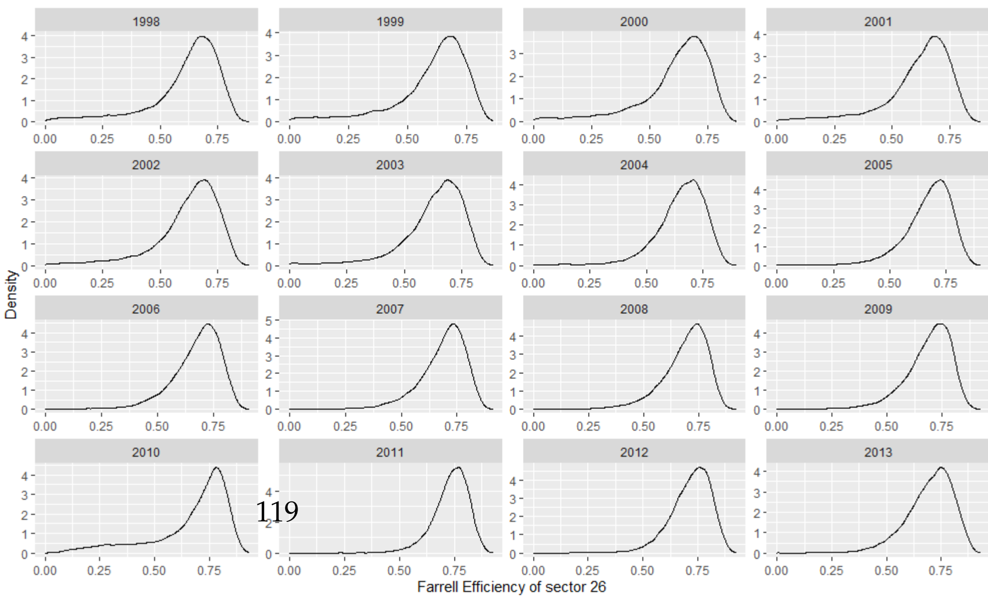


Figure 8.22: sector 26

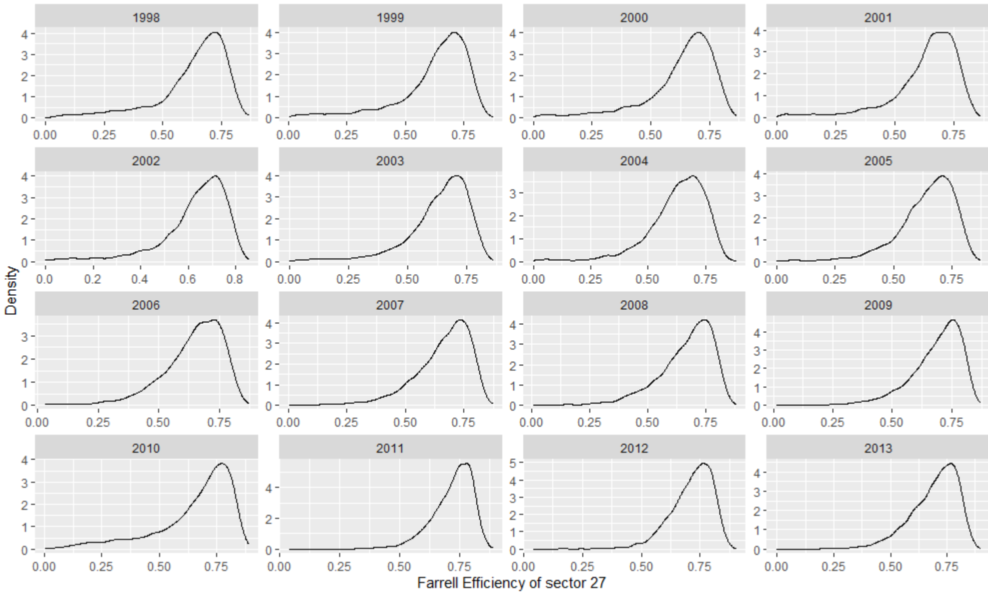


Figure 8.23: sector 27

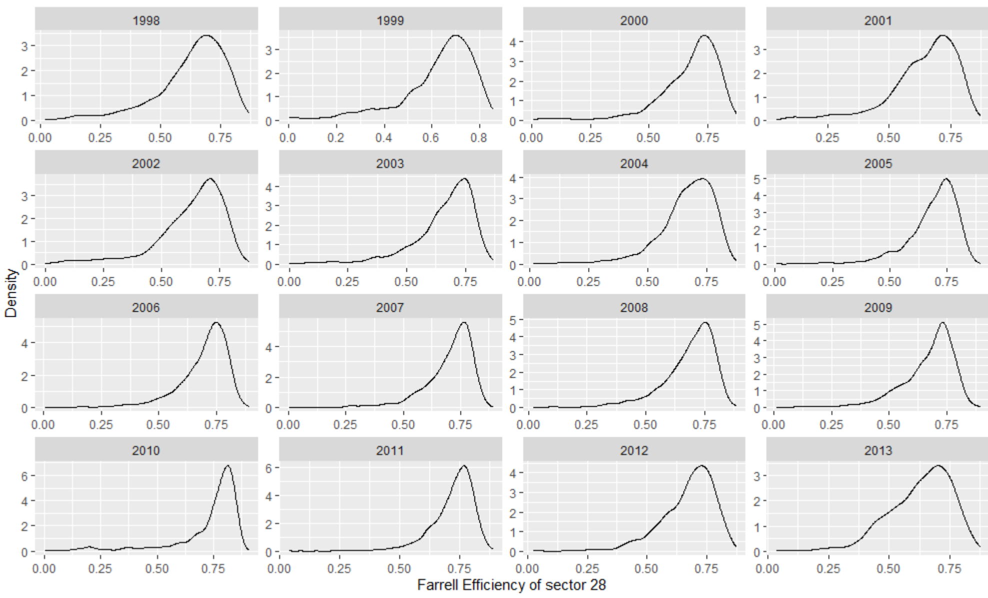


Figure 8.24: sector 28

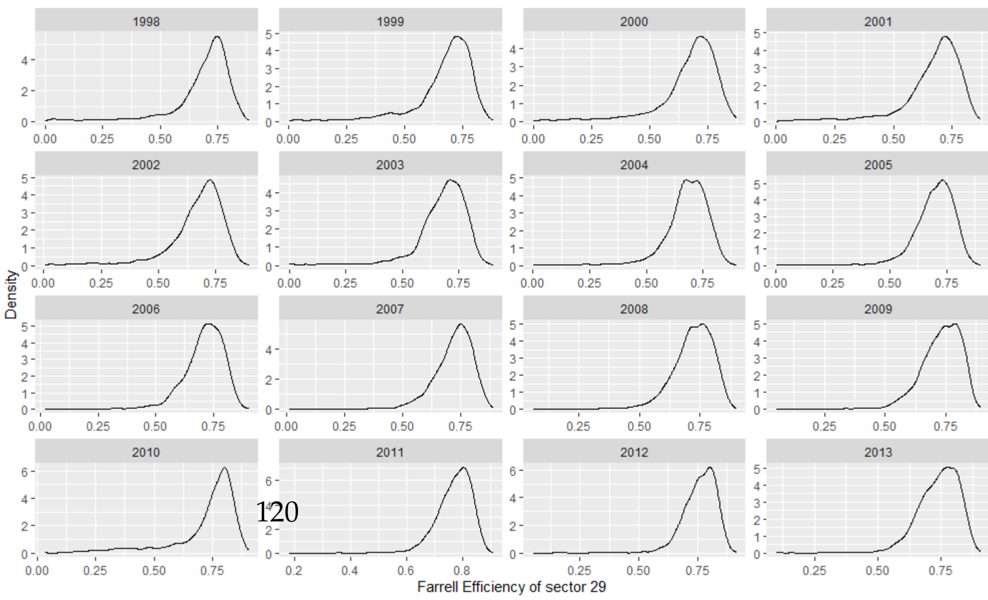


Figure 8.25: sector 29

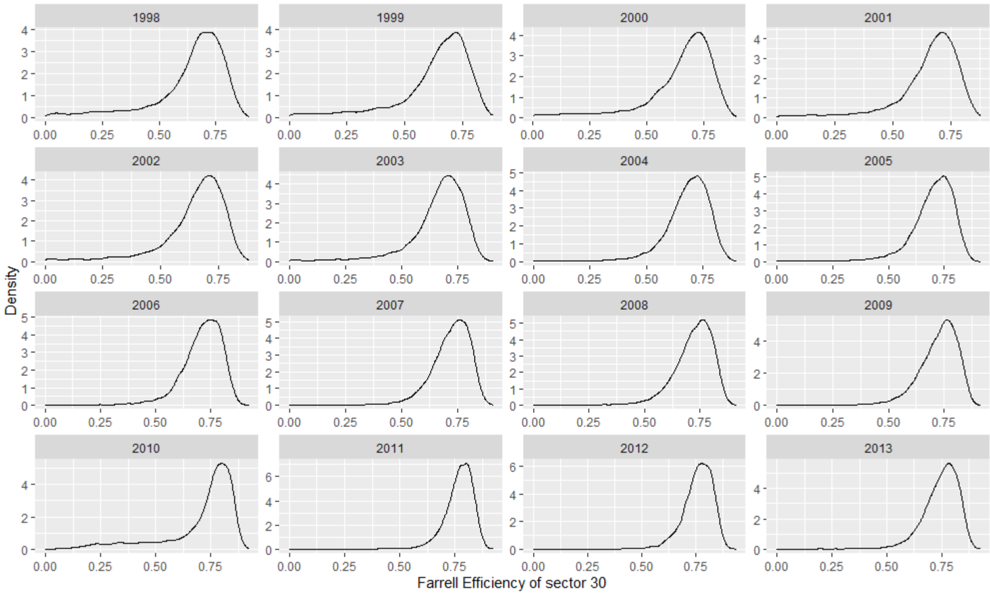


Figure 8.26: sector 30

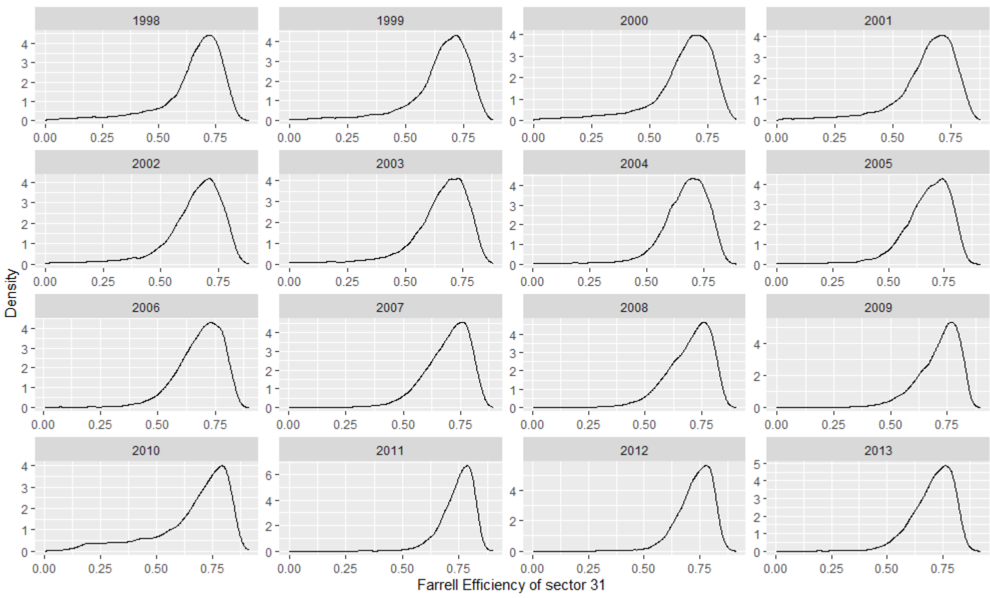


Figure 8.27: sector 31

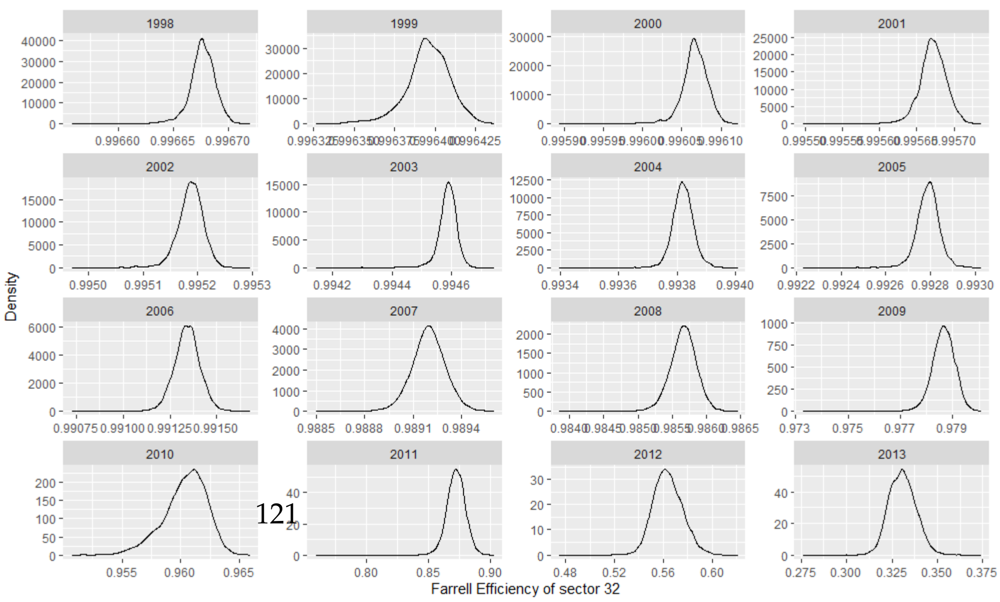


Figure 8.28: sector 32

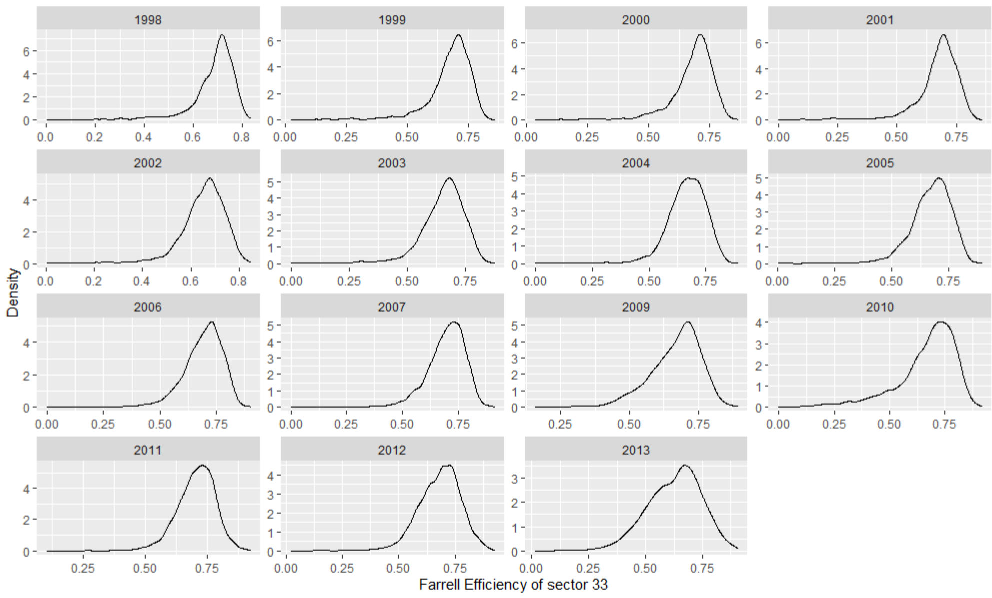


Figure 8.29: sector 33[30]

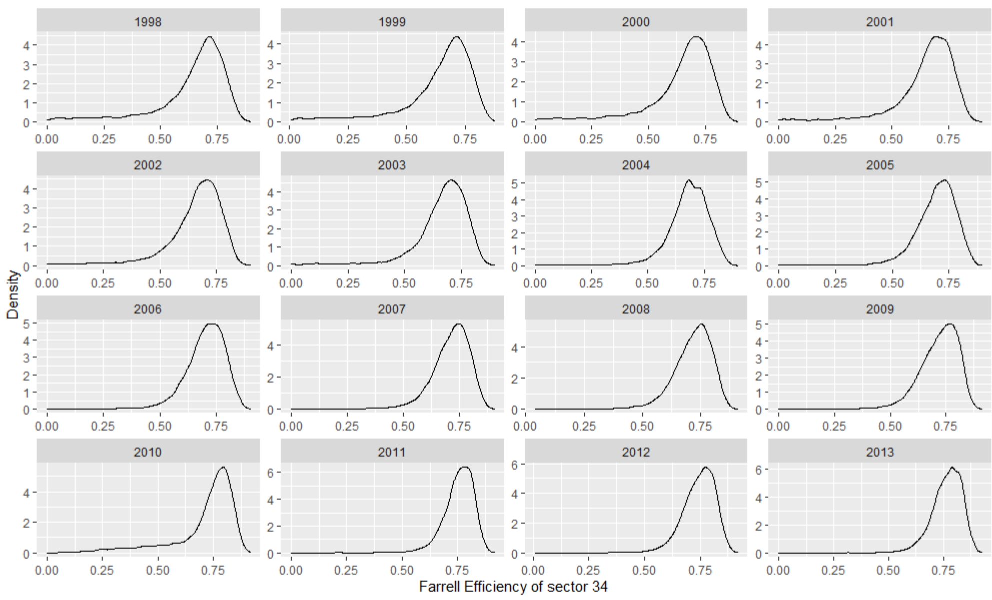


Figure 8.30: sector 34

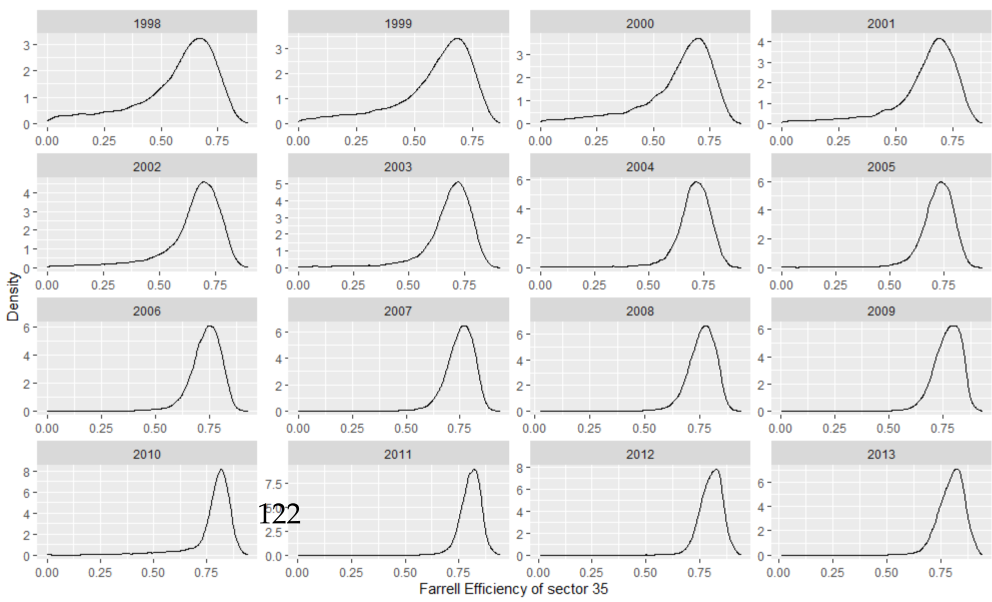


Figure 8.31: sector 35

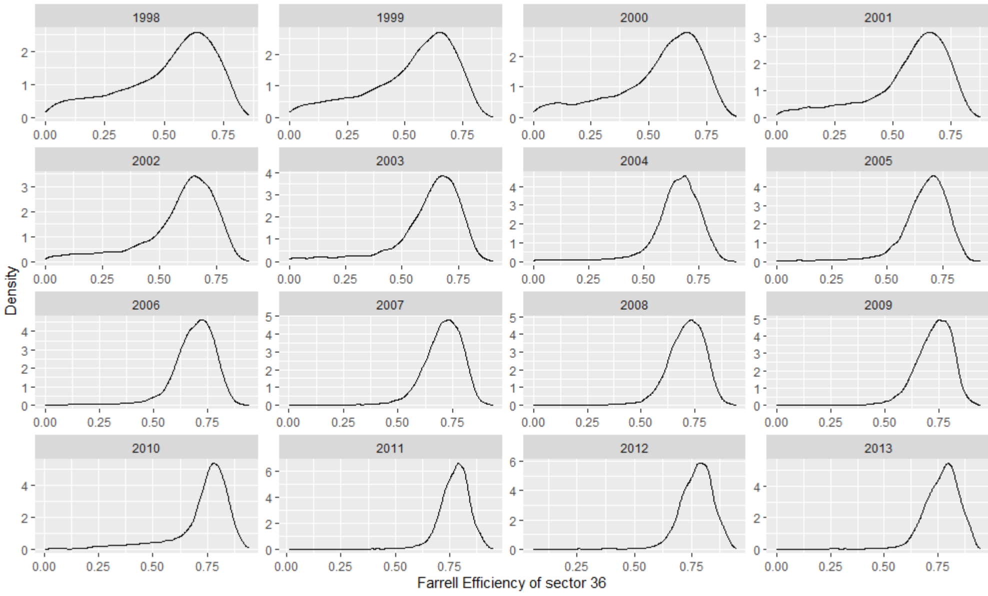


Figure 8.32: sector 36

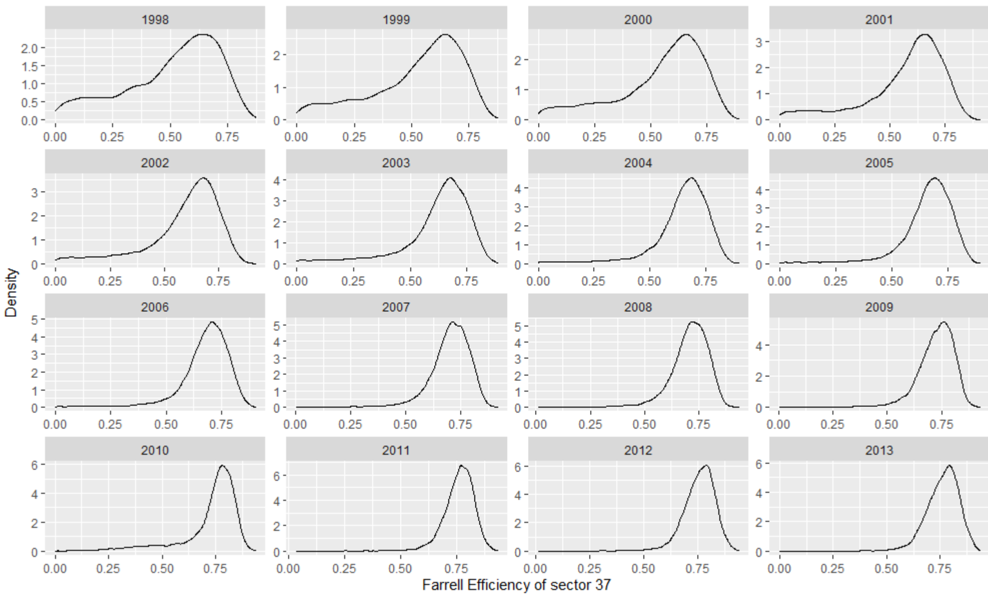


Figure 8.33: sector 37

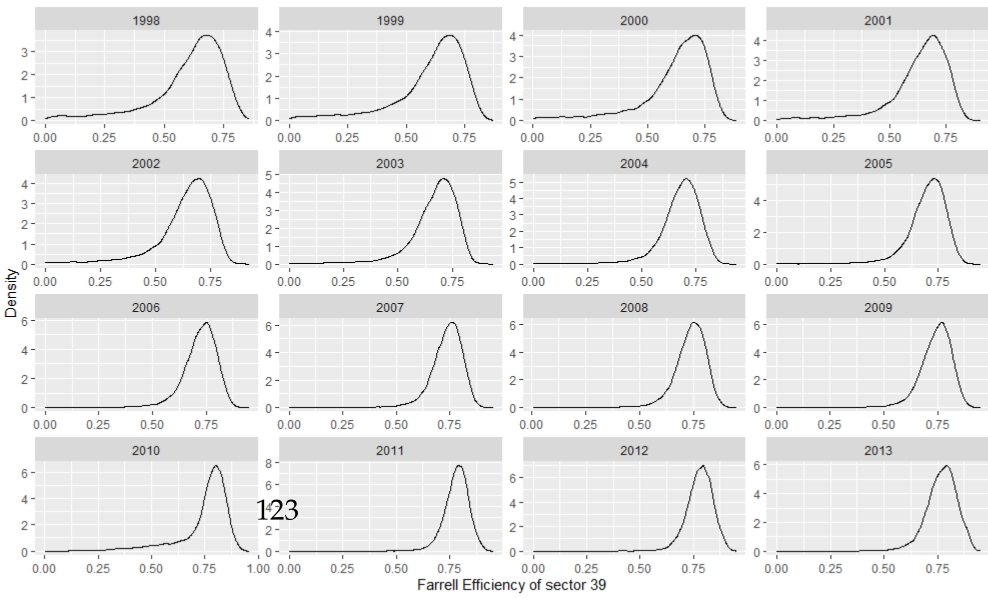


Figure 8.34: sector 39

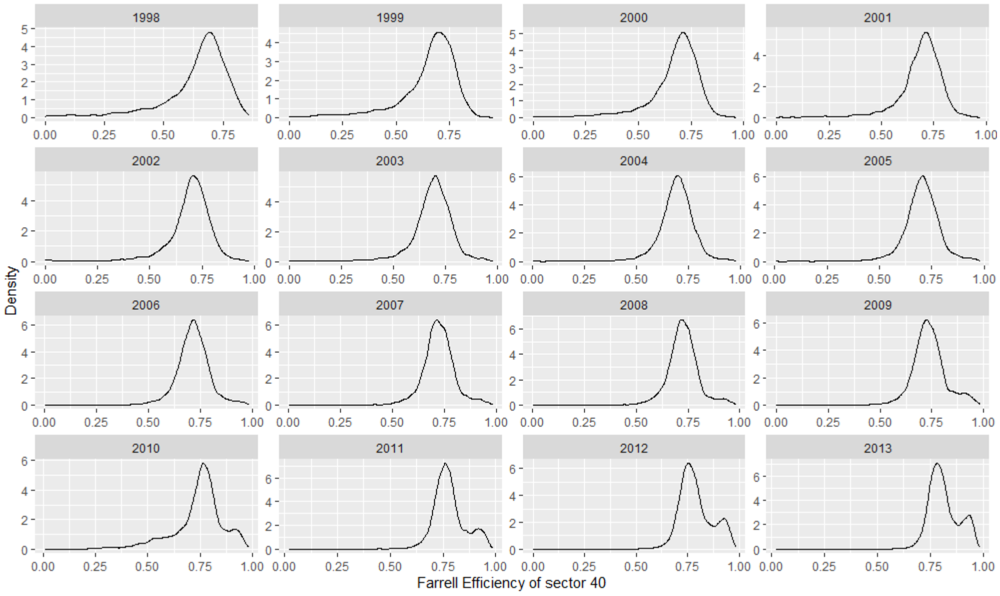


Figure 8.35: sector 40

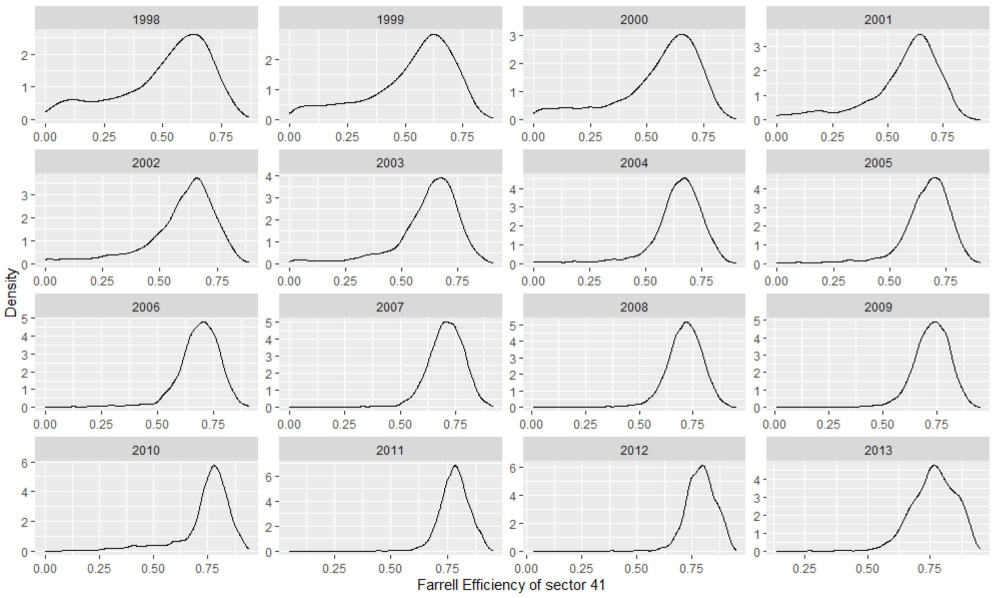


Figure 8.36: sector 41

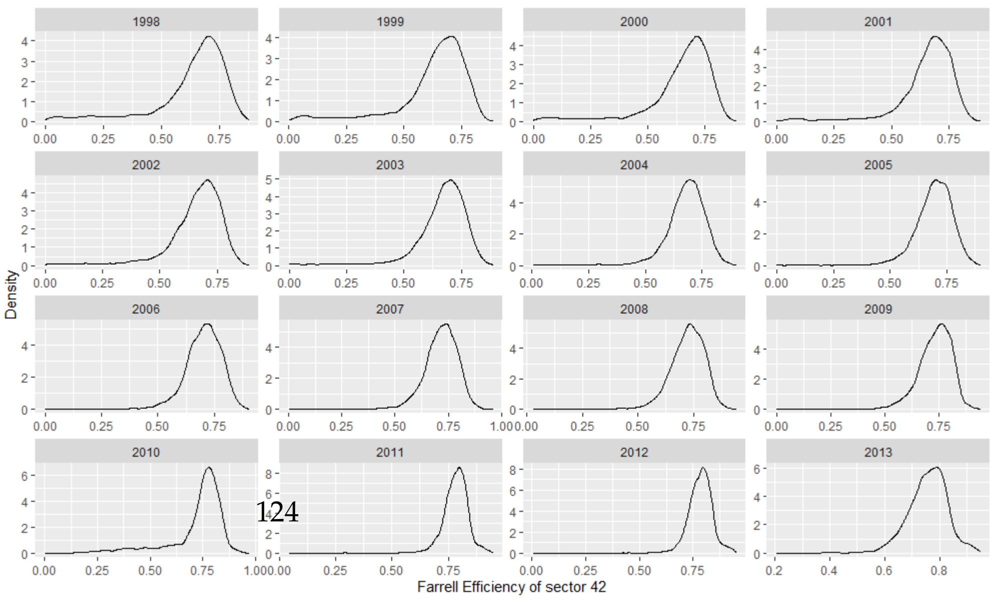


Figure 8.37: sector 42

8.5 Yearly shift of cumulative invent patents' marginal effect on Farrell efficiency

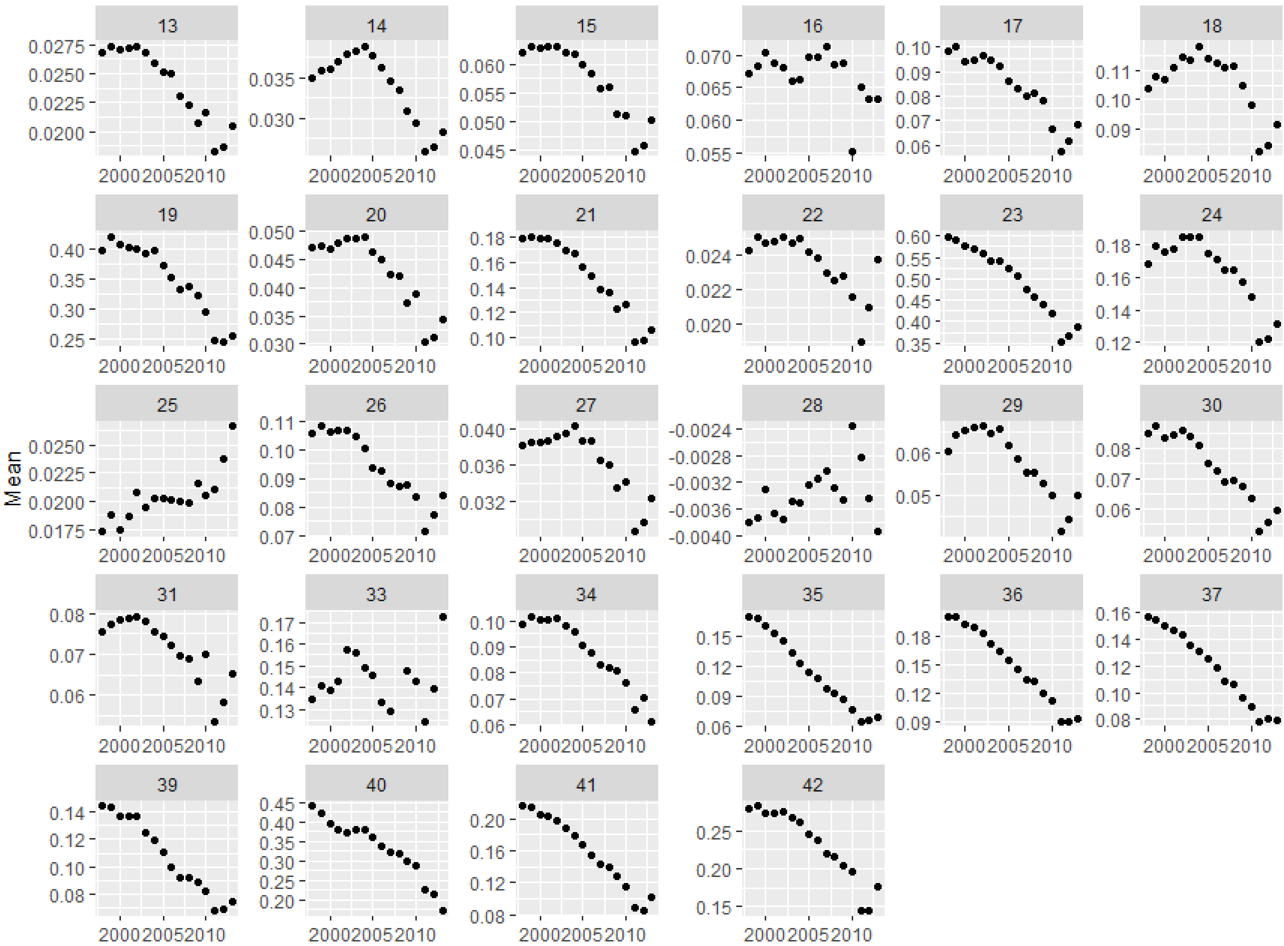


Figure 8.38: Yearly shift of cumulative invent patents' marginal effect on Farrell efficiency

8.6 Regression of BC95 Farrell efficiency on co-patenting networks

Table 8.7: Regression of BC95 Farrell efficiency on co-patenting networks

	Pooling	FE	RE	Arellano-vcovHC(FE)	FDGLS
(Intercept)	0.5466^{***}		0.5455^{***}		
	(0.0078)		(0.0106)		
Burt's constraint	-0.0562^{***}	-0.0547^{***}	-0.0593^{***}	-0.0547^{***}	-0.0489^{***}
	(0.0021)	(0.0031)	(0.0024)	(0.0056)	(0.0027)
betcent	0.8784	6.0395^{***}	4.8548^{***}	6.0395	3.2393
	(0.7835)	(1.4801)	(1.2448)	(6.4447)	(1.7606)
LogEmployee	0.0131^{***}	0.0003	0.0098^{***}	0.0003	-0.0065^{***}
	(0.0004)	(0.0008)	(0.0005)	(0.0014)	(0.0006)
localization	0.0028^{***}	0.0039^{**}	0.0030^{***}	0.0039	0.0020^{*}
	(0.0006)	(0.0012)	(0.0008)	(0.0022)	(0.0009)
urbanizationsize	0.0064^{***}	0.0206^{***}	0.0086^{***}	0.0206^{***}	0.0245^{***}
	(0.0007)	(0.0017)	(0.0010)	(0.0044)	(0.0015)
urbdiver	0.0081^{***}	-0.0082^{**}	0.0027	-0.0082^{*}	-0.0063^{**}
	(0.0012)	(0.0026)	(0.0016)	(0.0041)	(0.0019)
competition	-0.0004	-0.0004	-0.0008	-0.0004	-0.0001
	(0.0006)	(0.0009)	(0.0007)	(0.0013)	(0.0007)
StateOwned	-0.0222^{***}	0.0035	-0.0061^{**}	0.0035	-0.0028
	(0.0020)	(0.0021)	(0.0019)	(0.0033)	(0.0017)
HMT	0.0010	0.0085^{**}	0.0061^{*}	0.0085^{*}	0.0112^{***}
	(0.0025)	(0.0030)	(0.0026)	(0.0035)	(0.0023)
Foreign	0.0081^{***}	0.0103^{***}	0.0100^{***}	0.0103^{***}	0.0063^{***}
	(0.0021)	(0.0025)	(0.0022)	(0.0026)	(0.0018)
Collective	-0.0008	0.0029^{**}	0.0014	0.0029^{*}	-0.0013
	(0.0011)	(0.0011)	(0.0010)	(0.0013)	(0.0008)
Industry-Dummies	YES	YES	YES	YES	YES
R^2	0.2877	0.0589	0.6119		0.7081
Adj. R^2	0.2869	-0.2735	0.6115		
Num. obs.	38564	38564	38564		38564
s_idios			0.0709		
s_id			0.0638		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8.8: Hypothesis tests with respect to regression on co-patenting network(BC95)

Breusch-Pagan test for null hypothesis of homoskedasticity(FE)
BP = 19633, df = 39, p-value < 2.2e-16
Breusch-Godfrey/Wooldridge test for serial correlation in panel models(FE)
chisq = 75.353, df = 1, p-value < 2.2e-16
Breusch-Pagan LM test for cross-sectional dependence in panels(FE)
z = 210.44, p-value < 2.2e-16
Hausman Test with respect of regression on co-patenting network(FE vs. RE)
chisq = 616.89, df = 39, p-value < 2.2e-16

8.7 Regression of BC95 Farrell efficiency on citation networks

	Pooling	FE	RE	Arrleo-vcovHC	FEGLS	FGLS
(Intercept)	0.4853^{***}		0.5056^{***}			0.6036^{***}
	(0.0146)		(0.0196)			(0.0177)
authority	1.5039^{***}	0.7831^{***}	1.0734^{***}	0.7831^{**}	0.8188^{***}	1.6426^{***}
	(0.1442)	(0.1514)	(0.1417)	(0.2685)	(0.1645)	(0.1659)
hub	-0.1297^{***}	-0.0026	-0.0198	-0.0026	0.0470	-0.2495^{***}
	(0.0385)	(0.0427)	(0.0411)	(0.0632)	(0.0475)	(0.0422)
LogEmployee	0.0170^{***}	-0.0003	0.0119^{***}	-0.0003	-0.0069^{***}	0.0104^{***}
	(0.0005)	(0.0010)	(0.0007)	(0.0017)	(0.0007)	(0.0006)
localization	0.0016	0.0099^{***}	0.0048^{***}	0.0099^{***}	-0.0008	0.0001
	(0.0008)	(0.0016)	(0.0010)	(0.0027)	(0.0011)	(0.0009)
urbanizationsize	0.0074^{***}	0.0227^{***}	0.0075^{***}	0.0227^{***}	0.0180^{***}	0.0043^{***}
	(0.0011)	(0.0025)	(0.0015)	(0.0051)	(0.0017)	(0.0012)
urbdiver	0.0059^{***}	-0.0160^{***}	-0.0029	-0.0160^{**}	-0.0116^{***}	0.0006
	(0.0016)	(0.0033)	(0.0021)	(0.0058)	(0.0021)	(0.0019)
competition	-0.0026^{**}	-0.0029^{*}	-0.0029^{**}	-0.0029	-0.0032^{***}	-0.0035^{***}
	(0.0008)	(0.0012)	(0.0009)	(0.0020)	(0.0008)	(0.0008)
State-owned	-0.0194^{***}	-0.0021	-0.0091^{**}	-0.0021	0.0008	-0.0149^{***}
	(0.0032)	(0.0032)	(0.0029)	(0.0061)	(0.0022)	(0.0026)
HMT	-0.0269^{***}	-0.0001	-0.0097^{***}	-0.0001	0.0009	-0.0094^{***}
	(0.0026)	(0.0029)	(0.0025)	(0.0035)	(0.0020)	(0.0023)
Foreign	-0.0069^{**}	0.0013	-0.0028	0.0013	0.0005	-0.0024
	(0.0022)	(0.0025)	(0.0022)	(0.0028)	(0.0017)	(0.0019)
Collective	-0.0028	0.0025	0.0007	0.0025	-0.0028^{**}	-0.0031^{**}
	(0.0016)	(0.0015)	(0.0013)	(0.0017)	(0.0010)	(0.0012)
Industry-dummies	Yes	Yes	Yes	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8.9: Statistical models

Hypothesis tests with respect to regression on citation network(Farrell Efficiency)
Pesaran CD test for cross-sectional dependence in panels(FE)
z = 428.65, p-value < 2.2e-16
Breusch-Pagan test about homoscedasticity(FE)
BP = 7453.3, df = 39, p-value < 2.2e-16
Breusch-Godfrey/Wooldridge test for serial correlation in panel models(FE)
chisq = 224.64, df = 1, p-value < 2.2e-16
Hausman Test (FE vs. RE)
chisq = 578.46, df = 39, p-value < 2.2e-16
Hausman Test (FEGLS vs. FGLS)
chisq = 3009.4, df = 39, p-value < 2.2e-16

8.8 Regression of wagebenefit(logarithm) on co-patenting networks

Table 8.10: Regression of Wagebenefit per capita on co-patenting network

	Pooling	FE	RE	Arrleo-vcovHC	FDGLS
(Intercept)	-2.5196^{***}		-1.2502^{***}		
	(0.1552)		(0.1730)		
log(Burt's constraint + 1)	-0.3551^{***}	-0.3113^{***}	-0.3860^{***}	-0.3113^{***}	-0.3752^{***}
	(0.0327)	(0.0510)	(0.0374)	(0.0622)	(0.0464)
log(betcent + 1)	-30.8421^{***}	-9.9077	-25.4009^{*}	-9.9077	1.9815
	(7.2680)	(12.9883)	(10.8617)	(25.4416)	(12.1612)
log(localization + 1)	0.1858^{***}	0.0888	0.1526^{*}	0.0888	0.2174^{***}
	(0.0511)	(0.0929)	(0.0610)	(0.1129)	(0.0619)
log(urbanizationsize + 1)	1.2163^{***}	-0.0456	0.6190^{***}	-0.0456	-0.1824^{*}
	(0.0754)	(0.1131)	(0.0840)	(0.1594)	(0.0741)
log(urbdiver + 1)	-0.1526^{***}	-0.1070	-0.0819^{*}	-0.1070	0.0611
	(0.0295)	(0.0665)	(0.0372)	(0.0877)	(0.0605)
log(competition + 1)	0.0905^{***}	0.1301^{***}	0.1225^{***}	0.1301^{***}	0.1155^{***}
	(0.0175)	(0.0307)	(0.0205)	(0.0392)	(0.0265)
log(Output)	0.1833^{***}	0.2406^{***}	0.2044^{***}	0.2406^{***}	0.1892^{***}
	(0.0032)	(0.0074)	(0.0041)	(0.0128)	(0.0066)
log(BC95)	0.1656^{***}	0.0850^{**}	0.1275^{***}	0.0850	0.1901^{***}
	(0.0278)	(0.0323)	(0.0266)	(0.0566)	(0.0265)
log(CumPat+1)	0.0526^{***}	0.0595^{***}	0.0573^{***}	0.0595^{***}	0.0549^{***}
	(0.0043)	(0.0059)	(0.0045)	(0.0078)	(0.0051)
StateOwned	-0.0125	-0.1098^{***}	-0.0595^{***}	-0.1098^{***}	-0.1819^{***}
	(0.0170)	(0.0191)	(0.0167)	(0.0228)	(0.0145)
HMT	-0.0255	0.0367	0.0091	0.0367	0.0354
	(0.0221)	(0.0331)	(0.0246)	(0.0401)	(0.0290)
Foreign	0.0723^{***}	0.0807^{**}	0.0665^{**}	0.0807^{*}	0.0824^{***}
	(0.0189)	(0.0270)	(0.0205)	(0.0330)	(0.0237)
Collective	0.0052	0.0890^{***}	0.0263^{*}	0.0890^{***}	0.0726^{***}
	(0.0104)	(0.0129)	(0.0103)	(0.0159)	(0.0117)
Industry-Dummies	YES	YES	YES	YES	YES
R^2	0.2624	0.1430	0.4326		0.7593
Adj. R^2	0.2613	-0.3004	0.4318		
Num. obs.	28343	28343	28343		28343
s_idios			0.5160		
s_id			0.5206		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8.11: Hypothesis tests with respect to regression on co-patenting network(WageBenefit)

Pesaran CD test for cross-sectional dependence in panels(FE)
z = 521.94, p-value < 2.2e-16
Breusch-Pagan test about homoscedasticity(FE)
BP = 726.02, df = 40, p-value < 2.2e-16
Breusch-Godfrey/Wooldridge test for serial correlation in panel models(FE)
chisq = 387.63, df = 1, p-value < 2.2e-16
Hausman Test (FE vs. RE)
chisq = 401.3, df = 40, p-value < 2.2e-16

8.9 Regression of new product intensity on co-patenting networks

Table 8.12: Regression of new product intensity on copatenting network

	Pooling	FE	RE	Arrleo-vcovHC	FEGLS
(Intercept)	-0.5511^{***}		-0.2883^{**}		
	(0.0943)		(0.0913)		
log(Burt's constraint + 1)	-0.0121	-0.1131^{**}	-0.0776^{*}	-0.1131^{**}	-0.0211
	(0.0308)	(0.0374)	(0.0312)	(0.0417)	(0.0313)
log(betcent + 1)	-9.8004^{**}	-6.3941	-8.0020	-6.3941	-1.1603
	(3.3189)	(5.7115)	(4.6587)	(6.1664)	(5.5103)
log(localization + 1)	-0.0448	0.0223	-0.0160	0.0223	-0.0404
	(0.0350)	(0.0430)	(0.0352)	(0.0383)	(0.0339)
log(urbanizationsize + 1)	0.1472^{***}	-0.0301	0.0303	-0.0301	0.0143
	(0.0443)	(0.0448)	(0.0390)	(0.0332)	(0.0338)
log(urbdiver + 1)	0.1127^{***}	0.0836	0.1196^{***}	0.0836	0.1571^{***}
	(0.0214)	(0.0437)	(0.0263)	(0.0554)	(0.0375)
log(competition + 1)	-0.0172	-0.0079	-0.0168	-0.0079	0.0973^{***}
	(0.0133)	(0.0210)	(0.0146)	(0.0252)	(0.0175)
log(Output)	0.0182^{***}	0.0319^{***}	0.0209^{***}	0.0319^{***}	0.0206^{***}
	(0.0025)	(0.0051)	(0.0030)	(0.0066)	(0.0044)
log(BC95)	-0.0272	-0.0305	-0.0242	-0.0305	-0.0073
	(0.0220)	(0.0246)	(0.0201)	(0.0280)	(0.0194)
log(CumPat+1)	0.0357^{***}	0.0113^{**}	0.0193^{***}	0.0113^{*}	0.0141^{***}
	(0.0041)	(0.0043)	(0.0037)	(0.0051)	(0.0036)
StateOwned	-0.0211^{*}	-0.0051	-0.0118	-0.0051	-0.0110
	(0.0103)	(0.0102)	(0.0091)	(0.0115)	(0.0083)
HMT	-0.1105^{***}	-0.0118	-0.0817^{***}	-0.0118	-0.0111
	(0.0179)	(0.0260)	(0.0191)	(0.0287)	(0.0207)
Foreign	-0.0856^{***}	0.0902^{***}	-0.0116	0.0902^{*}	0.0735^{***}
	(0.0154)	(0.0249)	(0.0171)	(0.0418)	(0.0215)
Collective	-0.0493^{**}	0.0014	-0.0162	0.0014	0.0072
	(0.0157)	(0.0146)	(0.0134)	(0.0139)	(0.0120)
Industry-Dummies	YES	YES	YES	YES	YES
R^2	0.1189	0.0383	0.0612		0.8024
Adj. R^2	0.1133	-0.4931	0.0552		
Num. obs.	6444	6444	6444		6444
s_idios			0.1588		
s_id			0.2365		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Hypothesis tests with respect to regression on co-patenting network(New Product intensity)
Pesaran CD test for cross-sectional dependence in panels(FE)
z = 184.55, p-value < 2.2e-16
Breusch-Pagan test about homoscedasticity(FE)
BP = 721.24, df = 41, p-value < 2.2e-16
Breusch-Godfrey/Wooldridge test for serial correlation in panel models(FE)
chisq = 25.587, df = 1, p-value = 4.229e-07
Hausman Test (FE vs. RE)
chisq = 140.58, df = 38, p-value = 1.089e-13

8.10 Regression of patent count on co-patenting network

Return code 1: gradient close to zero (gradtol)
Log-Likelihood: -2173.884 Num.obs=3493 t=2
6 free parameters
Estimates:

	Estimate	Std. error	t value	Pr(> t)
log(R & D exp_lg)	0.025991	0.002096	12.399	< 2e-16 ***
log(R & D exp)	-0.01109	0.004051	-2.738	0.006190 **
log(Burt's constraint + 1)	-3.94607	0.244902	-16.113	< 2e-16 ***
log(betcent + 1)	15.8341	4.206933	3.764	0.000167 ***
log(Output + 1)	0.505342	0.046589	10.847	< 2e-16 ***
Year2007_dummy	0.035797	0.020702	1.729	0.083782 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bibliography

[1] Cohen W M, Levinthal D A. Absorptive capacity: A new perspective on learning and innovation.[J].*Administrative science quarterly*, 1990: 128-152.
[2] Zaheer A, Bell G G, Benefiting from network position: firm capabilities, structural holes, and performance.[J]. *Strategic management journal*, 2005, 26(9): 809-825.
[3] Belderbos R, Cassiman B, Faems D, et al., Co-ownership of intellectual property: Exploring the value-appropriation and value-creation implications of co-patenting with different partners.[J].*Research policy*, 2014, 43(5): 841-852.
[4] Sammarra A, Biggiero L. Heterogeneity and specificity of InterFirm knowledge flows in innovation networks.[J]. *Journal of management studies*, 2008, 45(4): 800-829.
[5] Iino T, Inoue H, Saito Y U, et al. How does the global network of research collaboration affect the quality of innovation?.[J].*The Japanese Economic Review*, 2021, 72: 5-48.
[6] Aktamov S, Zhao Y. Impact of network centrality positions on innovation performance of the firm: evidence from china automobile industry.[J].*Business Management and Strategy*, 2014, 5(1): 164.
[7] Cohen W M, Levinthal D A. Innovation and learning: the two faces of R & D.[J]. *The economic journal*, 1989, 99(397): 569-596.
[8] Gilsing V, Nooteboom B, Vanhaverbeke W, et al. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density.[J]. *Research policy*, 2008, 37(10): 1717-1731.
[9] Brondino G. Productivity growth and structural change in china (1995 – 2009): A subsystems analysis.[J]. *Structural Change and Economic Dynamics*, 2019, 49: 183-191.
[10] Cucculelli M, Ermini B. Risk attitude, product innovation, and firm growth. Evidence from Italian manufacturing firms.[J].*Economics Letters*, 2013, 118(2): 275-279.
[11] Vasudeva G, Zaheer A, Hernandez E. The embeddedness of networks: Institutions, structural holes, and innovativeness in the fuel cell industry.[J]. *Organization science*, 2013, 24(3): 645-663.
[12] Story V M, Boso N, Cadogan J W. The form of relationship between firmlevel product innovativeness and new product performance in developed and emerging markets.[J]. *Journal of Product Innovation Management*, 2015, 32(1): 45-64.

[13] Holmes T J, Stevens J J. Geographic concentration and establishment scale.$[J]$. *Review of Economics and Statistics*, 2002, 84(4): 682-690.
[14] Nakamura R, Paul C J M. Measuring agglomeration.$[M]$.*Handbook of regional growth and development theories*. Edward Elgar Publishing, 2019.
[15] Tian Z, Gottlieb P D, Goetz S J. Measuring industry co-location across county borders.$[J]$. *Spatial Economic Analysis*, 2020, 15(1): 92-113.
[16] Beyene J, Moineddin R. Methods for confidence interval estimation of a ratio parameter with application to location quotients.$[J]$. *BMC Medical Research Methodology*, 2005, 5(1): 1-7.
[17] Moineddin R, Beyene J, Boyle E. On the location quotient confidence interval.$[J]$. *Geographical analysis*, 2003, 35(3): 249-256.
[18] Duranton G, Overman H G. Testing for localization using micro-geographic data.$[J]$. *The Review of Economic Studies*, 2005, 72(4): 1077-1106.
[19] Strotebeck F. The Location Quotient – Assembly and application of methodological enhancements.$[Online]$. Available at: https://www.mpra.ub.uni-muenchen.de.html
[20] Hackman S T. Production economics: integrating the microeconomic and engineering perspectives.$[J]$. *Springer Science & Business Media*, 2007.
[21] Hu T, Zhang Y. A spatial – temporal network analysis of patent transfers from US universities to firms.$[J]$. *Scientometrics*, 2021, 126(1): 27-54.
[22] Lin H L, Li H Y, Yang C H. Agglomeration and productivity: Firm-level evidence from China's textile industry.$[J]$. *China Economic Review*, 2011, 22(3): 313-329.
[23] Zeng J, Liu D, Yi H. Agglomeration, Structural Embeddedness, and Enterprises' Innovation Performance: An Empirical Study of Wuhan Biopharmaceutical Industrial Cluster Network.$[J]$. *Sustainability*, 2019, 11(14): 3922.
[24] Mori T, Smith T E. Analysis of industrial agglomeration patterns: an application to manufacturing industries in Japan.$[J]$. *KIER Discussion Paper*, 2011, 794.
[25] Ahuja G. Collaboration networks, structural holes, and innovation: A longitudinal study.$[J]$. *Administrative science quarterly*, 2000, 45(3): 425-455.
[26] Nakamura R. Contributions of local agglomeration to productivity: Stochastic frontier estimations from Japanese manufacturing firm data.$[J]$. *Papers in Regional Science*, 2012, 91(3): 569-597.
[27] Brandt L, Van Biesebroeck J, Zhang Y. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing.$[J]$. *Journal of development economics*, 2012, 97(2): 339-351.
[28] Guimarães P, Figueiredo O, Woodward D. Dartboard tests for the location quotient.$[J]$. *Regional Science and Urban Economics*, 2009, 39(3): 360-364.
[29] De Beule F, Van Beveren I. Does firm agglomeration drive product innovation and renewal? An application for Belgium.$[J]$. *Tijdschrift voor economische en sociale geografie*, 2012, 103(4): 457-472.
[30] Gagliardi L. Employment and technological change: on the geography of labour market adjustments.$[Online]$. Available at: http://eprints.lse.ac.uk/64499/.html
[31] Guan J, Liu N. Exploitative and exploratory innovations in knowledge network and collaboration network:

A patent analysis in the technological field of nano-energy.[J]. *Research policy*, 2016, 45(1): 97-112.
[32] Holmes T J, Stevens J J. Geographic concentration and establishment scale.[J]. *Review of Economics and Statistics*, 2002, 84(4): 682-690.
[33] Newsom J T. Distinguishing between random and fixed: Variables, effects, and coefficients. [$Online$]. Available at: http:www.upa.pdx.edu/IOA/newsom/mlrclass/ho_ randfixd.doc/.html (2006-07-19), 2006.
[34] Li D, Lu Y, Wu M. Industrial agglomeration and firm size: Evidence from China.[J]. *Regional Science and Urban Economics*, 2012, 42(1-2): 135-143.
[35] Feldman M P, Audretsch D B. Innovation in cities:: Science-based diversity, specialization and localized competition.[J]. *European economic review*, 1999, 43(2): 409-429.
[36] Yun J H J, Jeong E S, Park J S. Network analysis of open innovation.[J]. *Sustainability*, 2016, 8(8): 729.
[37] Tan J, Zhang H, Wang L. Network closure or structural hole? The conditioning effects of network – level social capital on innovation performance.[J]. *Entrepreneurship Theory and Practice*, 2015, 39(5): 1189-1212.
[38] O'Donoghue D, Gleave B. A note on methods for measuring industrial agglomeration.[J]. *Regional studies*, 2004, 38(4): 419-427.
[39] Cho T S, Shih H Y. Patent citation network analysis of core and emerging technologies in Taiwan: 1997 – 2008.[J]. *Scientometrics*, 2011, 89(3): 795-811.
[40] Rasmussen S. *SProduction economics: the basic theory of production optimisation*.[$Book$]. Springer Science & Business Media, 2012.
[41] Sebestyén T, Varga A. Research productivity and the quality of interregional knowledge networks.[J]. *The Annals of Regional Science*, 2013, 51(1): 155-189.
[42] Martin P, Mayer T, Mayneris F. Spatial concentration and plant-level productivity in France.[J]. *Journal of Urban Economics*, 2011, 69(2): 182-195.
[43] Zheng D, Hao S, Sun C, et al. Spatial correlation and convergence analysis of eco-efficiency in china.[J]. *Sustainability*, 2019, 11(9): 2490.
[44] Millo G, Piras G. splm: Spatial panel data models in R.[J]. *Journal of statistical software*, 2012, 47: 1-38.
[45] Bai J, Li S, Wang N, et al. Spatial spillover effect of new energy development on economic growth in developing areas of China—An empirical test based on the spatial Durbin model.[J]. *Sustainability*, 2020, 12(8): 3249.
[46] Tenaye A. Technical efficiency of smallholder agriculture in developing countries: The case of Ethiopia.[J]. *Economies*, 2020, 8(2): 34.
[47] Rawat P S, Sharma S. TFP growth, technical efficiency and catch-up dynamics: Evidence from Indian manufacturing.[J]. *Economic Modelling*, 2021, 103: 105622.
[48] Crescenzi R, Rodríguez-Pose A, Storper M. The territorial dynamics of innovation in China and India.[J]. *Journal of economic geography*, 2012, 12(5): 1055-1085.
[49] Yang G C, Li G, Li C Y, et al. Using the comprehensive patent citation network (CPC) to evaluate patent value.[J]. *Scientometrics*, 2015, 105(3): 1319-1346.
[50] Barnum D T, Gleason J M. Using Panel Data Analysis to Estimate Confidence Intervals for the DEA Efficiency of Individual Urban Paratransit Agencies.[J]. *Applied Economics* 2009.

[51] Zhang H. How does agglomeration promote the product innovation of Chinese firms?.[J]. *China economic review*, 2015, 35: 105-120.
[52] Tian Z. Measuring agglomeration using the standardized location quotient with a bootstrap method.[J]. *Journal of Regional Analysis & Policy*, 2013, 43(2): 186-197.
[53] Zhengge T U, Renjun S. Productivity growth in China's large industrial firms: patterns, causes, and implications.[J]. *Journal of Chinese Studies*, 2016, 1(1): 1-16.
[54] Z Tian - Measuring agglomeration using the standardized location quotient with a bootstrap method.[J]. *Journal of Regional Analysis* & Policy, 2013
[55] Andre Lemelin, Fernando Rubiera-Morollón & Ana Gómez-Loscos, Measuring Urban Agglomeration: A Refoundation of the Mean City-Population Size Index.[J], *Social Indicators Research* volume 125, pages589 – 612 (2016)
[56] JM Wooldridge, [$Book$].Econometric analysis of cross section and panel data, 2010
[57] J LeSage, RK Pace, Introduction to spatial econometrics,2009
[58] H Kelejian, G Piras, [$Book$].Spatial econometrics, 2017
[59] WW Powell, KW Koput, L Smith-Doerr, Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology [J], *Administrative science quarterly*, 1996
[60] Gautam Ahuja, Collaboration networks, structural holes, and innovation: A longitudinal study.[J]. *Administrative science quarterly*,First Published September 1, 2000
[61] A Zaheer, GG Bel, Benefiting from network position: firm capabilities, structural holes, and performance. [J]. *Strategic management journal*, 2005
[62] Philipp Herzog Jens Leker, Open and closed innovation- different innovation cultures of different strategies. [J]. *Journal of technology management* 52(3-4): 322-343
[63] Strotebeck, Falk, The Location Quotient – Assembly and application of methodological enhancements. [$Online$]. Available at: mpra.ub.uni-muenchen.de.html
[64] TJ Holmes, JJ Stevens, Geographic concentration and establishment scale. [J]. *Review of Economics and Statistics*, 2002
[65] R Moineddin, J Beyene, E Boyle, On the location quotient confidence interval. [J]. *Geographical analysis*,16 November 2010
[66] P Guimarães, O Figueiredo, D Woodward. 2009, Dartboard tests for the location quotient. [J]. *Regional Science and Urban Economics*, Volume 39, Issue 3, May 2009, Pages 360-364
[67] B.W. Silverman.1998. [$Book$]. Density Estimation for Statistics and Data Analysis(1st Edition)
[68] Markus Eberhardt, Christian Helmers and Zhihong Yu, Is the Dragon Learning to Fly? An Analysis of the Chinese Patent Explosion. The University of Nottingham Research Paper No. 2011/16
[69] JM Kleinberg. Authoritative sources in a hyperlinked environment.[J]. *Journal of the ACM (JACM)*,Volume 46Issue 5Sept. 1999 pp 604 – 632
[70] Jianwei Dang, Kazuyuki Motohashi. Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality.[J].*China Economic Review*, Volume 35, September 2015, Pages 137-155.

[71] Yuyu Chen, Mitsuru Igami, Mo Xiao. Privatization and Innovation: Productivity, New Products, and Patents in China. November 17, 2015 Preliminary draft prepared for the NBER Productivity Semina

[72] M Arellano. Computing Robust Standard Errors for Withingroups Estimators. $[J]$. *Oxford bulletin of Economics and Statistics*, 1987- ideas.repec.org

[73] BH Baltagi. $[Book]$. Econometric Analysis of Panel Data, 2008- Springer° [74] Olsen, Jakob Vesterlund and Arne Henningsen, Investment utilization and farm efficiency in Danish agriculture. FOI working paper 2011/13, Institute of Food and Resource Economics,University of Copenhagen

[75] Figueiredo, O.; Guimaraes, P; Woodward, D. Localization Economies and Establishment Scale: A Dartboard Approach, FEP Working Papers, Nr. 247, September 2007.

[76] Zi-Lin He, Tony W. Tong, Yuchen Zhang. A database linking Chinese patents to China's census firms. $[J]$.*Scientific Data* 5, Article number: 180042 (2018)

[77] Z Xie, X Zhang. The patterns of patents in China. $[J]$.*China Economic Journal*,Pages 122-142,2s015

[78] M. Hashem Pesaran. General Diagnostic Tests for Cross Section Dependence in Panels.$[J]$.*Empirical Essconomics*,60, pages13 – 50 (2021)

[79] P. A. P. Moran. Notes on continuous stochastic phenomena. $[J]$.*Biometrika* Vol. 37, No. 1/2 (Jun., 1950), pp. 17-23 (7 pages)