

学位論文及び審査結果の要旨

横浜国立大学

氏名	Baojun Fang
学位の種類	博士（経済学）
学位記番号	国府博甲第64号
学位授与年月日	令和4年3月24日
学位授与の根拠	学位規則（昭和28年4月1日文部省令第9号）第4条第1項及び横浜国立大学学位規則第5条第1項
研究科(学府)・専攻名	国際社会科学府経済学専攻
学位論文題目	The Chinese urban economy and ASIE manufacture firms' activity: an empirical research based on the micro data (中国都市部経済と ASIE 製造業企業の活動：個票データに基づく実証研究)
論文審査委員	主査 横浜国立大学 大森 義明 教授 横浜国立大学 奥村 綱雄 教授 横浜国立大学 藤生 源子 教授 横浜国立大学 張 馨元 准教授 新潟県立大学 秋山 太郎 教授

論文の要旨

Preparation of data

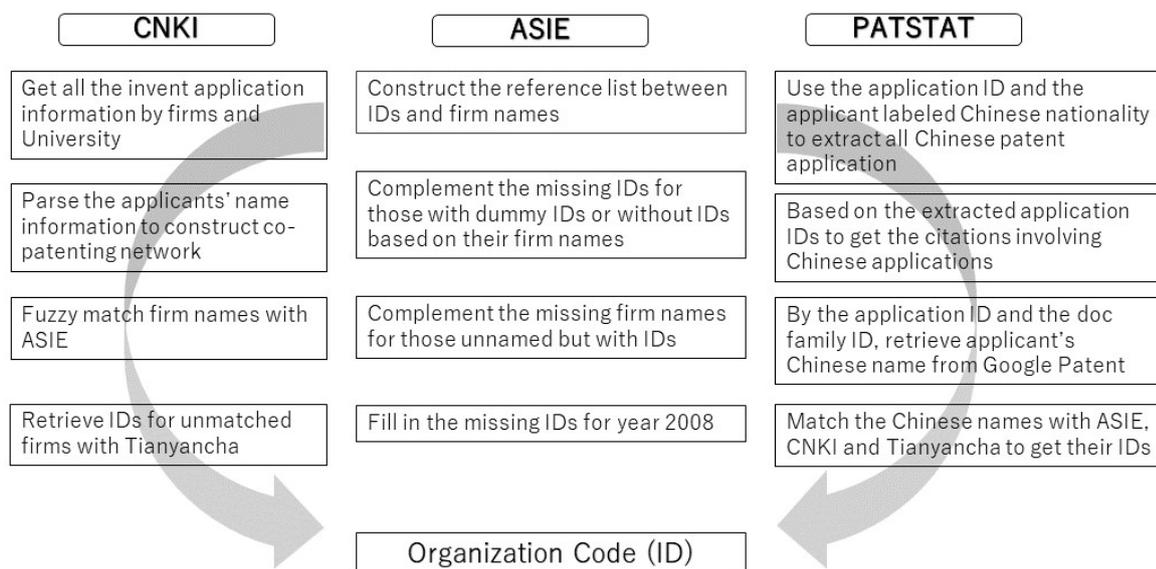
According to Chinese statistics bureau, ASIE(Annual survey of industrial enterprises)' s involves all the state-owned enterprises and the non-state-owned enterprises with turnover over 5 million Chinese dollars (20 million Chinese dollars since 2011). And the industries can be divided into 3 subgroups of mining, manufacturing and public services(water, electricity and gas).By now the available database for public is based year 1998-2013. Due to the authority and continuity of this database's composition by statistics bureau, 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:

- 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.

- 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. And to certify the code a certification rule was defined by official regulation GB 11714-1997 and another amended version in 2007.
- ④ Price deflation. As the responses by sample firms are based on values at the responding time point, the price level is not unified. Different from Loren Brandt (2011), I use the yearly composite PPI (i.e., producer price index for industrial products) published by NBS with year 1998 as base level 100.
- ⑤ Difficulty to link with other data source about the Chinese manufacturing firms such patent and import-export database.

I adopted new data sources and computer science technique to get a more specific, precise, and higher-matching ratio linkage result, which is of higher quality compared with existing linkage. Moreover, based on the derived copatenting and citation networks, it gives every researcher of interest a more accessible and reliable source to probe into the production mechanism and innovation-incentive mechanism with graph theory in addition to the traditional economics analytic theories.

Specific process to link different database can is as bellow:



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.

Based on existing data entries in ASIE, one verifiable organization code can be linked to one or multiple names but not vice versa. 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 business registry website. 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.

Copatenting data source here is based on the published 1985-2013 patent database by CNKI. 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. 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.

After the data wrangling and parsing, fuzzy matching with ASIE and inquiry into Tianyancha are carried out. Except those firms only existing before organization code system was implemented, nearly all firms are attached with their IDs. And among the total count of 914,591 invention patents, 782,227 can be matched with ASIE enlisted firms which amounts to 70,420. Within the constructed co-patenting network, 6,910 out of 22,795 net nodes are ASIE enlisted firms.

Citation data source here contains PATSTAT Version 5.15 and Google patent. And the network constructed 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 patents is not unordinary. And there are 316,785 patents embedded in this citation network out of which 115,456 are applied by 17,684 Chinese firms. Moreover 8,382 are ASIE enlisted firms.

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

The discrete agglomeration indices are very charming for policymakers, 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 a discrete agglomeration index awkward to cope with the agglomeration phenomenon across different administrative regions.

Yearly localization quotient is 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 (Duranton, et al. 2004) 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 based on the zip codes attached to ASIE.

Traditional regression models mostly stipulate the assumptions about exogeneity 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.

FE model (i.e., fixed effect model), RE model (i.e., seemingly unrelated regression with presumed error structure), fixed effect generalized least squares model (FEGLS) and feasible general least squares model (FGLS) are adapted here as the basic model. In addition, the main spatial panel models adopted here include three categories which impose restrictions on the interaction parameters ρ , θ , λ , and the included models include: 1. SAC (spatial autoregressive combined model) 2. SDEM (Spatial Durbin error model) . Based a spatial lag two stage least squares estimator by H.Kelejian (2004) on the these two spatial models the spillover effect can be inferred by the direct impacts and indirect impacts.

Influence of R & D network on firm's activity

Network related statistics here includes degree of node, betweenness centrality and burt's constraint measure and hyperlink-Induced topic search (HITS) algorithms.

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 refer 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, citation and copatenting networks are 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 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. Therefore, 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 if 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 economic 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 estimated.

Different from the social network analysis embedded in the co-patenting network, authority and hub scores can act as the proxy of innovativeness and absorptiveness bibliometrically.

Hypothesis 7: Authority scores are positively related to the production efficiency.

Hypothesis 8: Hub scores are positively related to the production efficiency.

Main findings: Burt's constraint is negatively related with the technical efficiency and the new

product intensity. It can be inferred that opportunistic firms are more efficient to produce with same mercenary or physical inputs. One the side, firms with lower constraint are more innovative with launching new products. As to result with respect to regression on wage, constraints have negative effects on the wages.

Contrast with constraints, betweenness centrality is not significantly related to any of the dependent variables except the output of patent application. And different with hub, authority is solely significant in positive relation with the efficiency. It is consistent with classic economic theory that firms with high-level technology boast higher productivity.

審査結果の要旨

2章は、共同特許と引用ネットワークを含む本論文で用いる分析データを構築する。

中国統計局の企業センサス ASIE のデータは、調査対象企業の脱落、2008年以降の企業IDの欠落やそれ以前との不一致、他のデータとの連携の難しさなどのデータの研究利用上の問題がある。このため、ASIEを利用した実証研究は、Loren Brandt, et al. (2011)に従って、1998年～2007年までのデータを用いて行われてきた。Fang氏は、企業IDと企業名・住所等の情報の組み合わせた利用、コンピューター・サイエンスの手法を用いた企業の同定、中国のビジネスレジストリ情報サイトである天眼查の情報との照合、最終的な手作業による確認などを複合的に行った上で、従来は研究上利用ができなかった2008年～2013年のデータを含む品質、精度が高いデータベース化を行っている。さらに、中国データベースCNKの1985年から2013年の特許データを用い、特許の出願者の企業をASIEの企業とマッチングさせた上で、共同出願ネットワークを構築する。また、中国国内の特許データベースは引用関連のデータを含まないため、主としてPATSTATに含まれる引用関連データを用い、引用ネットワークデータを構築している。

3章前半は、中国の製造業の地理的分布を分析する。離散的集積指数であるLQ指数の修正版、連続グローバル局所化指数・分散指数を求め、比較する。離散的集積指数は、地理的単位をまたがる集積を把握できず、地理的単位の統合・分割によって影響を受ける。この問題が生じない指標として、空間を連続的に扱い、企業間の距離の分布をカーネル推定し、企業の立地がランダムに行われる場合についてのモンテカルロシミュレーションから求められる信頼区間との差に基づいて集積を識別するグローバル局所化・分散の指数(Durantón, et al. 2004)を計算している。産業別の結果の比較から、グローバル局所化・分散指数からは、集積の状況が安定している産業が多く、離散集積指数は集積の進行を過大評価すること等がわかる。3章後半は集積と企業規模との関係を実証分析する。マーシャルの外部性は地域内の集積と企業の平均規模の間に正の相関をもたらすと理論は予測する。固定効果、変量効果、固定効果一般化最小二乗、実行可能な一般最小二乗の各モデル、さらに2種類の空間パネルモデルを推定する。30の産業毎に集積と企業規模との関係の分析から、推定法によらず、集積は企業の平均規模に正で有意な効果を持つことがわかる。

4章は、特許共同出願ネットワークと特許引用ネットワークがイノベーションや生産効率性といった企業パフォーマンスに与える効果を分析する。ネットワークは異なる主体が持つ知識の組み合

わせを促進し、イノベーションを促進し、企業の効率性を改善すると考えられる。特許の共同出願ネットワークを共同研究開発ネットワークの代理変数として、また、特許引用ネットワークを企業間の知識フローのネットワークの代理変数として用いる。企業のネットワーク上の位置を表す統計量として、共同出願ネットワークでは媒介中心性、バートの拘束度、方向を持つネットワークではである特許引用ネットワークでは、ハイパーリンク誘導トピック検索 (HITS) アルゴリズムの権威値、ハブ値を用いる。企業のパフォーマンスの尺度として ASIE データの新製品売上比率、労働者 1 人当たり賃金、生産の効率性 (Farrell 効率性) を用いる。実証結果は、拘束度が生産の効率性、新製品の売り上げ、賃金のいずれに対しても負の効果を持つことを示す。すなわち、共同研究ネットワークで緊密に結ばれていない部分を橋渡しする企業は、優れたパフォーマンスを挙げていることが示唆される。一方、媒介中心性は生産の効率性、新製品の売り上げ、賃金のいずれに対しても有意な効果を持たない。また、権威値は生産の効率性に対して有意な正の効果を持ち、優れた知識の発信元である革新性が高い企業は、生産効率性が高いことが示唆される。既存研究の多くは、イノベーションのパフォーマンス尺度として、企業の特許の被引用数を用いる。しかし、特許引用は真のイノベーションの経済的成果ではなく、引用を成果の指標とすると、特許引用ネットワークのイノベーションへの効果も分析できない。Fang 氏は、ASIE のユニークなデータである新製品売上をイノベーションの経済的成果の尺度として、また、ASIE から計算された企業の生産効率性等を企業のパフォーマンスの尺度として用いている点でも優れている。

以上より、本論文審査委員一同は、本学府の博士号審査基準②に照らして、Baojun Fang 氏の学位請求論文「The Chinese urban economy and ASIE manufacture firms' activity : an empirical research based on the micro data」が博士 (経済学) の学位を授与するに値するものとして、判断する。

令和 4 年 1 月 7 日

審査委員主査 横浜国立大学大学院国際社会科学研究院教授 大森義明

審査委員 横浜国立大学大学院国際社会科学研究院教授 奥村綱雄

審査委員 横浜国立大学大学院国際社会科学研究院教授 藤生源子

審査委員 横浜国立大学大学院国際社会科学研究院准教授 張馨元

審査委員 新潟県立大学国際経済学部教授 秋山太郎

参考 : Baojun Fang 氏の指導委員会の構成員は以下の通りである。

責任指導教員 横浜国立大学大学院国際社会科学研究院教授 大森義明

指導教員 横浜国立大学大学院国際社会科学研究院教授 奥村綱雄

指導教員 横浜国立大学大学院国際社会科学研究院教授 藤生源子