Sources of Relatedness between Industries and Revealed Comparative Advantage: Evidence from China*

Yao Amber Li[†]
HKUST

Linke Zhu[‡]
SUFE

This Version: August 2014
First Draft: June 2014

Abstract

This paper presents empirical evidence from disaggregated Chinese industrial data (1998-2009) to uncover *how* the revealed comparative advantage (RCA) of an industry in a region is affected by the relatedness between industries. To identify different sources of relatedness between industries, we quantify relatedness by employing the density measure developed by Hidalgo, Klinger, Barabási and Hausmann (*Science*, 2007). We further decompose the sources of relatedness into four mechanisms: input linkage, output linkage, labor market pooling, and knowledge spillovers. To address the identification issue, we employ the FDI policy variations across industries in China to further test the mechanisms of industry relatedness and their impacts on RCA.

There are three main findings. First, all four sources of relatedness significantly increase the future probability of an industry to have RCA in that region, and the future RCA of an industry is mostly driven by its relatedness with other local industries regarding labor pooling and knowledge spillovers. Second, industries with higher capital or skill intensity, and thus, industrial upgrading, benefit more from being located in the same region with other industries using similar labor, while benefit less from knowledge spillover from other local industries. In other words, the effect of knowledge spillover for capital intensive industry is much smaller than that for labor intensive industries. Third, the aforementioned findings are further confirmed by the tests using FDI policies in China: FDI encouragement policies tend to be more successful in regions with higher labor pooling density, but less successful in regions with higher knowledge spillover density. Our results are robust to alternative measures of labor pooling mechanism, different aggregation level of regions, and controlling for endowment driven comparative advantage.

Keywords: relatedness, revealed comparative advantage, FDI, industrial upgrading, industrial dynamics, China, input-output, labor market pooling, knowledge spillover

---

*The authors thank Tony Venables, John Romalis, Albert Park, Sandra Poncet, Yi Lu, and Liuqiu Zhao for helpful discussions on this project. Yao Amber Li gratefully acknowledges financial support from the Institute for Emerging Market Studies (Project No.: IEMS14BM04). We also thank Yiye Cui, Yicong Xia, Chen Kan, and Feng Zhou for excellent research assistance during this project. All errors are our own.

†Li: Department of Economics and Faculty Associate of the Institute for Emerging Market Studies (IEMS), Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR-PRC. Email: yaoli@ust.hk. URL: ihome.ust.hk/∼yaoli. Tel: (852)2358 7605; Fax: (852)2358 2084. Research Affiliate of the China Research and Policy Group at University of Western Ontario.

‡Zhu: School of International Business Administration, Shanghai University of Finance and Economics. Email: zhulinke83@gmail.com. URL: https://sites.google.com/site/econlinkezhu/.

1 Introduction

The evolution of the revealed comparative advantage (RCA) of an industry in a region as a dynamic process is of great importance to economic growth, international trade, and regional development: it is not only interesting to economists and policy makers, but also relevant to private investors in their business decisions. As is well known, international trade theory emphasizes comparative advantage in explaining the distribution of industries across countries or regions, and comparative advantages are usually determined by technology, endowments, or scale economies. However, unfortunately, compared with other traditional determinants of comparative advantages, we know much less about *how* the RCA is affected by relatedness between industries. In other words, little has been explored about how the linkages between an industry and the existing industrial structure of a region determine the evolution of the RCA of that industry in that region.[1] The objective of this paper is to fill a gap in the literature by linking the evolution of RCA to various channels of relatedness between industries and by empirically assessing the importance of each channel.

We start with identifying the relatedness between industries by employing an industry-region-time specific density measure developed by Hausmann and Klinger (2007) and then used in Hidalgo et al. (2007) and Bahar et al. (2014). In this emerging literature, the seminal work of Hausmann and Klinger (2007) and Hidalgo et al. (2007) develops the concept of "product space" and use co-exporting patterns of products across countries to create the density measure of relatedness between products. In other words, the density of a product proxies for the existence of other exports that share similar technologies or inputs, as measured by their co-occurrence across countries (Bahar et al., 2014). Their findings show that countries tend to develop new products that are most closely related to the products which they already have comparative advantage in. However, the underlying mechanisms of relatedness has not yet been fully explored. It has also been acknowledged that the results could be driven by a common third factor that escapes the controls of the existing set-up (Bahar et al., 2014). Our paper extends this literature by investigating different sources of the relatedness between industries and their impacts on RCA as well as by pursuing an endeavor in identification strategies to uncover causality.

We first conduct a similar exercise as in Hidalgo et al. (2007) by regressing forward RCA on existing RCA and an overall density measure of industry relatedness. In this exercise we build our overall relatedness measure upon coagglomeration patterns of industries across Chinese regions instead

[1] Note that New Economic Geography (NEG) literature normally explores only one channel of relatedness between industries, namely, the input-output linkages, and most studies in this literature are theoretical work (Krugman and Venables, 1996; Venables, 1999; Amiti, 2005; Baldwin and Venables, 2011; Puga and Venables, 1996, e.g.).

of co-export of goods.[2] The results show that industry relatedness significantly and positively affect RCA. We further decompose the sources of relatedness into four Marshallian mechanisms – input linkage, output linkage, labor market pooling, and knowledge spillovers – inspired by the economic geography literature (e.g., Marshall, 1920 ; Ellison et al., 2010, among others). To examine how the relations between relatedness and RCA depend on industry heterogeneity, we also include H-O factors and categorize industries by their capital and skill intensities.

To alleviate the identification issue (e.g., causality and omitted variables), we employ FDI policy variations across industries in China (i.e., either encouragement or restriction at industry level) to further test the four key mechanisms of relatedness and their impacts on the evolution of RCA. Since FDI policy variations are the arguably exogenous shocks imposed on Chinese industries, we use interactions between FDI policies and four Marshallian based density measures of relatedness to check whether the same FDI policy has different effects on an industry's RCA in different regions given different existing industrial structures in those regions. This is analogous to a difference-in-difference strategy. In other words, we expect that the FDI encouragement policy for an industry is more successful in a region if that industry is better embedded in the current industrial structure of that region through the key mechanisms of relatedness between industries.

Our empirical analysis relies on data extracted from four sources. First, industry(-region)-level measures of the characteristics of Chinese industries is constructed by several databases obtained from the National Bureau of Statistics of China (NBS). The regional unit used in our analysis is defined at either province level or city level. Second, the co-production pattern of multi-industry firms across industries are based on the export data from China's General Administration of Customs. Third, labor market pooling measures using occupation data as robustness is computed from the General Social Survey of China.[3] Lastly, the FDI policy variables is constructed from the Catalogue for the Guidance of Foreign Direct Investment Industries issued by China's National Development and Reform Commission (NDRC) and Ministry of Commerce in various years. Our sample period is 1998-2009.

We test the determinants of industry relatedness and its impact on evolution of RCA as well as industrial upgrading with Chinese data because China presents a good setting to examine those issues: China is a fast-growing, export-oriented economy; China is also a large developing country, both geographically and in market size. Although the literature has shown a substantial amount of data regarding the process of structural transformation (or, more recently, industrial dynamics) in today's

[2]Later we will use co-export and co-production as proxy for proximities between industries due to knowledge spillovers.
[3]We also use education data from the NBS to compute the labor market pooling effect.

advanced economies, there are still very limited studies, if none, investigating the underlying driving forces at disaggregated industry level in developing countries (Herrendorf et al., 2014). Therefore, using Chinese data presents distinctive advantages for our study.

There are three main findings. First, the general co-agglomeration based relatedness and all four Marshallian based sources of relatedness significantly increase the future probability of an industry to have RCA in that region. The future RCA of an industry is mostly driven by the relatedness regarding labor pooling and knowledge spillovers.

Second, industries with higher capital intensities or skill intensities, and thus, industrial upgrading, benefit more from being located in the same region with other industries using similar labor, while benefit less from knowledge spillover from other local industries. In other words, the effect of knowledge spillover for capital intensive industry is much smaller than that for labor intensive industries. One possible explanation for these results is that more capital or skill intensive industries normally use better-quality workers and thus may benefit more from improved employer-employee matching resulting from a larger regional labor pool. Meanwhile, these industries may benefit less from knowledge spillovers since they normally use better technology than other industries. Thus, the expectation that other firms in related industries may learn from them and the concern of knowledge leakage may deter new entrants of these industries.

Third, the aforementioned findings are further confirmed by the tests using FDI policies in China: FDI encouragement policies tend to be more successful in regions with higher labor pooling density, but less successful in regions with higher knowledge spillover density. Our results are robust to alternative measures of labor pooling mechanism and controlling for regional-level endowment driven comparative advantages.

The key value-added of this paper is to provide compelling empirical analysis that the evolution of RCA and industrial upgrading depend on various Marshallian based sources of inter-industry relatedness, and meanwhile feature industrial heterogeneity regarding factor intensities. Moreover, this paper employs the arguably exogenous industry-level FDI policy shocks in China to carefully identify the causes of industry relatedness and their impacts on RCA at regional level, and thus contributes to the emerging literature that features network analysis and uses density measures to explore country-level comparative advantages.

This paper also contributes to several branches of the literature on industrial dynamics, comparative advantage, and economic geography. In the first two branches of the literature, relatedness between industries is either neglected (e.g., Herrendorf et al., 2013; Ju et al., 2013, among others),[4] or presents its importance while lacks the analysis of different sources of underlying mechanisms for relatedness (e.g., Hausmann and Klinger, 2007; Hidalgo et al., 2007; Poncet and de Waldema, 2013).

Our analysis of decomposing sources of inter-industry relatedness into various mechanisms is inspired by the literature of economic geography and New Economic Geography (NEG). However, our paper differs from this literature in at least two aspects. First, the economic geography literature usually focuses on Marshallian forces with no attention to the endowment driven comparative advantages (e.g. Ellison et al., 2010; Dumais et al., 2010). Second, the NEG literature introduces comparative advantages, but the development of industries in one region is determined primarily by the input-output linkages between industries and industry factor intensities (e.g. Puga and Venables, 1996; Venables, 1999; Amiti, 2005; Baldwin and Venables, 2011). Thus, other Marshallian forces are often neglected. In addition, there are very few empirical studies in the NEG literature and the existing ones primarily focus on the determinants of industrial location (e.g. Midelfart-Knarvik et al., 2001; Cabral et al., 2013; Bao et al., 2013).

The remainder of the paper is organized as follows. Section 2 describes different data sources we use. Section 3 presents measurement of industry relatedness and other variables of interest. Section 4 displays the econometric specifications and reports main findings as well as robustness checks. The last section concludes.

2 Data

Our analysis of the relatedness between industries and industrial upgrading relies on data extracted from four sources. First, industry(-region)-level measures of the characteristics of Chinese industries is constructed by several databases obtained from the National Bureau of Statistics of China (NBS). Second, co-integration data is computed upon the trade data from China's General Administration of Customs. Third, labor market pooling measures using occupation data is computed from the General Social Survey of China. Lastly, the FDI policy variables is constructed from the Catalogue for the Guidance of Foreign Direct Investment Industries issued by China's National Development and Reform Commission (NDRC) and Ministry of Commerce in various years. We briefly discuss the construction of our dataset in turn as follows.

[4] Also see Herrendorf et al. (2014) for a comprehensive review.

2.1 Industry(-Region)-Level Production and Industrial Characteristics Data

To build industry(-region)-level production and industrial characteristics data, we combine two firm-level databases with the input-output sector level data. All these databases are obtained from NBS.

The main database we use is the Annual Survey of Chinese Industrial Enterprises (CIE), conducted and maintained by the NBS. Our sample period is 1998-2009. The CIE database covers all state-owned enterprises (SOEs), and above-scale non-state-owned enterprises (with annual sales of at least 5 million RMB).[5] This database has been widely used by previous studies of Chinese economy (e.g., Brandt et al., 2012, among many others). It contains detailed firm-level information of manufacturing enterprises in China, such as the location of the firm (province, city, or county level), employment, capital stock, gross output, value added, wage payment, firm identification (e.g., company name, zip code, contact person, etc.), and many financial variables of the firm. Between 1998 and 2009, the approximate number of firms covered in the CIE data varied from 165,000 to 321,000. The number of firms increased over time, mainly because manufacturing firms in China have been growing rapidly over the sample period and more firms reached the threshold for inclusion. This CIE database contains 29 2-digit manufacturing industries and 425 adjusted 4-digit industries.[6] We will use the most disaggregated industry classification afforded by the data to construct industry relatedness measures. The most useful information in the CIE data is the location of each firm, the gross output, employment, and the industry code. We aggregate the gross industrial output and employment by province-industry-year triplet to construct the basic sample for our analysis.

The second database we use is the 2004 Chinese Economic Census data which contain all-scale industrial enterprises to supplement the CIE data. The reason we need the 2004 census data is because the census data provide detailed educational compositions of workers for each firm such that we are able to compute the skill intensity at industry level and the similarity between industries regarding their education levels. There are five education categories: junior high school or below, senior high school, diploma, undergraduate/college, and postgraduate. We also use capital stock (measured by fixed assets) from the 2004 census data to compute capital intensity together with skill intensity to control for industrial characteristics. Note that the 2004 census data contain more firms than the CIE data. To keep consistency over the sample, we use firm identification code to match the 2004 CIE

[5] It equals US$732,257 approximately, according to the official end-of-period exchange rate in 2009, reported by the central bank of China.

[6] Note that China adopted a new classification system for Chinese industrial classification (CIC) codes (GB/T 4754-2002) to replace the old classification system (GB/T 4754-1994) that had been used from 1995 to 2002. To make the CIC codes consistent across the whole sample period, we used the concordance between the two sets of CIC codes provided by Brandt et al. (2012) and eventually we obtained 425 adjusted 4-digit industries.

data with the 2004 census data, and only keep those firms who are present in both databases in 2004.

Last but not least, to measure the input-output relationships between industries, we use the 2007 Input-Output Tables published by the NBS. The tables provide input-output relationships between 135 IO sectors. We focus on manufacturing which correspond to 2-digit IO sector code 11 (Corn Milling) to 91 (Recycling and Disposal of Waste) and we further drop sector 91 which is the recycling industry. Then we use the concordance between the I-O sectors and the CIC industries provided by the NBS to match each 4-digit CIC industry from the CIE and the census data to the I-O sectors.[7] Eventually we mapped 416 4-digit adjusted CIC industries into 80 I-O sectors. This aggregation step is conducted because we need the input-output relations to compute the measures of industrial relatedness that require to take the vertical linkages between industries into account. To be more precise and consistent, we prefer to conduct all industry-level analysis at the I-O sector level. Therefore, all our baseline analyses in this paper are carried out based on these 80 IO sectors.

2.2 Chinese Customs Data

China's General Administration of Customs provides us with the universe of all Chinese trade transactions by importing and by exporting firm at the HS 8-digit level. We use this transactional-level Chinese Customs data mainly for examining the co-production patterns among multi-industry firms to infer the knowledge spillovers between industries (see Section 4.3 for more details). To ensure the exported products are indeed produced by those firms, we exclude all intermediary firms from the customs data, following the similar method as in Ahn et al. (2011) and Tang and Zhang (2012) by finding the presence of phrases (such as "trading", "exporting", and "importing") in their company names.[8] Then we use the firm-level exports data for year 2000 which is the earliest year when the Customs data is available for us and aggregate firm exports to 2-digit IO sector level according to the concordance between the I-O sectors and the HS codes issued by the NBS.

[7]The concordance between 4-digit CIC industry and 2-digit IO sector is relatively good. There are 425 adjusted 4-digit CIC industries that correspond to the 80 IO sectors we focus on. Only 9 of these CIC industries correspond to more than one IO sector. We simply drop these 9 industries whenever we have to aggregate data from 4-digit CIC industry to 2-digit IO sector. In our sample period, these 9 industries account for about 2 percent of total manufacturing employment.

[8]As company names in the Customs Database are written in Chinese, we search for "mao yi", "wai mao", "wai jing", "jin chu kou", "jing mao", "gong mao", and "ke mao" in firm names.

2.3 Occupation Data

The occupation data from the 2003 Chinese General Social Survey (GSS) is used to construct measures of labor market pooling between industries based on occupational similarity. This survey is an annual or biannual questionnaire survey of China's urban and rural households aiming to monitor systematically the changing social structure in urban and rural China. The GSS database is constructed by the Survey Research Center at Hong Kong University of Science and Technology and the Department of Sociology at Renmin University of China. The 2003 GSS is the earliest data available and contains occupational information. Each respondent in the survey reported all her past occupations and we assign the most recently reported occupation as the final occupation in 2003 to that respondent. In the original GSS data, there are 272 occupations in total which correspond to all 92 2-digit Chinese industries (including service industries and those in the primary sector). After dropping data in primary and tertiary sectors, we obtained 168 occupations distributed across 29 2-digit CIC industries in manufacturing which will be used to construct occupational similarities for labor market pooling measure in future.

2.4 FDI Policy Data

A final source of data are FDI policy variables that are constructed from the Chinese government lists that specifies which industries are encouraged, restricted, or prohibited for foreign investment. This list, provided in the Catalogue for the Guidance of Foreign Investment Industries, was first published in 1995 and was revised subsequently in 1997, 2002, 2004, and 2007 by China's National Development and Reform Commission (NDRC). These FDI policy changes present a unique opportunity for us to test the effect of relaxing FDI restrictions on industrial upgrading and its relationship with industry relatedness. Thus, we construct two FDI policy index variables, "FDI encouragement" and "FDI restriction", at 4-digit CIC industry level based on the detailed contents in the Catalogue.[9]

To construct these two FDI index variables, we manually search the keyword in the explanatory notes for Chinese National Economy Classification System (GB/T4754-2002) at 4-digit CIC industry level based on the detailed contents in the Catalogue. If a CIC industry contains some products or processes which are encouraged in the Catalogue, the industry is classified as "encouraged" industry (FDI encouragement=1); if a CIC industry contains some products or processes which are restricted

[9] A similar index has been first developed by Blonigen and Ma (2010), which examines the effect of FDI regulations on the composition of Chinese exports, and further adopted by Sheng and Yang (2013), which investigates the effects of these FDI policies on the export product varieties.

or even prohibited in the Catalogue, it is classified as "restricted" industry (FDI restriction=1). Note that a few industries contain both encouraged and restricted products, then those industries will be called hybrid industries and receive zero value for policy variables. The timing of policy change is employed in the coding and thus we obtain two time variant FDI policy variables. Finally, we aggregate the FDI policy variables into I-O sectors for future analysis.

3 Measurement of Main Variables

We expect that the revealed comparative advantage (RCA) of an industry in a region is affected by its relatedness with the existing industrial structure of that region. We employ the *density* measure created by Hidalgo et al. (2007) to quantify this relatedness[10]:

$$\omega_{it}^{k} = \sum_{j} x_{jt}^{k} \phi_{ij} \tag{1}$$

where ω_{it}^{k} is the density around industry i for region k in year t, $x_{jt}^{k} = 1$ if $RCA_{jt}^{k} > 1$ and 0 otherwise, and ϕ_{ij} is the matrix of relatedness between industries i and j. Note that it is assumed that ϕ_{ij} do not change over time. We define RCA as:

$$RCA_{jt}^{k} = \frac{Emp_{jt}^{k} / \sum_{j} Emp_{jt}^{k}}{\sum_{k} Emp_{jt}^{k} / \sum_{k} \sum_{j} Emp_{jt}^{k}} \tag{2}$$

where Emp_{jt}^{k} is the employment of industry j in region k in year t. The *density* measure defined above will be the base of the construction of our key explanatory variables in our empirical analysis.

Hidalgo et al. (2007) exploit the co-export pattern of goods across countries to measure relatedness (i.e., if two goods are observed to have high probability of being co-exported by the same country, they are regarded to have high proximity). We follow their methodology, but exploit coagglomeration patterns of industries across Chinese regions to measure overall relatedness between industries. The disadvantage of this measure is that it does not distinguish between different sources of relatedness between industries. To address this problem, we follow Ellison et al. (2010) to use three different measures of relatedness, each corresponding to a specific source of relatedness between industries. We

[10] We drop the denominator $\sum_{j} \phi_{ij}$ in their original definition as we believe that it only creates measurement error. For example, if $\phi_{1j} = 0.01\phi_{2j}$ for all j, then the industrial structure in any region will have the same effect on industry 1 and 2 according to their original definition, while one should expect that industry 1 is much more affected in this case since it has much stronger relatedness with other industries.

describe the construction of each of these measures in the following subsections.

3.1 Overall relatedness based on coagglomeration patterns

To replicate the results of Hidalgo et al. (2007), we construct an overall measure of relatedness between industries using the following equation:

$$Relate_{ij} = \min(Prob(RCA_i^k > 1|RCA_j^k > 1), Prob(RCA_j^k > 1|RCA_i^k > 1)) \tag{3}$$

We compute $Relate_{ij}$ both at province-level and city-level using the CIE data in year 1998. The idea behind this measure is that if two industries are observed to coagglomerate in one region (in the sense that both have RCA in that region) with high probability, then they are more likely to have high relatedness with each other. The advantage of this measure is its comprehensiveness, while the disadvantage is that it says nothing about the sources of the relatedness. A further disadvantage is that this measure is highly endogeneous when used to explain RCA dynamics across regions since it is based on RCA measures itself. Thus we construct the following more detailed and more exogeneous measures of relatedness between industries.

3.2 Input-output relationship

To reduce trade costs, firms tend to locate near their suppliers or customers. As a result, regions are more likely to gain or maintain RCA in industries that have closer input-output relationships with their existing RCA industries. To measure input-output relationships between industries, we use the 2007 Input-Output Tables published by the National Bureau of Statistics of China (NBS). The tables provide input-output relationships between 135 IO sectors. We focus on manufacturing which correspond to IO sector 11 to 91 and we further drop sector 91 which is the recycling industry and sector 87 due to data limitation. All our analyses in this paper are carried out for the remaining 79 I-O sectors.

We define $Input_{ij}$ as the share of industry is inputs that come from industry j (the total "inputs" here also include factors that are used in production). We also define $Output_{ij}$ as the share of industry i's outputs that are sold to industry j.

3.3 Labor market pooling

Our second measure of relatedness is based on the similarity between industries in terms of types of labor used. We expect that industries are more likely to grow in a region where they can find a large pool of labor that are suitable for them to use. A large labor pool can reduce costs for various reasons including risk sharing and better matching (for a summary of these theories, see Ellison et al. (2010)). Apparently, firms are more likely to find such a labor pool in regions where the predominant industries use similar types of labor as they do. We measure the extent to which industries use similar types of labor through the educational employment patterns across industries. This information is available from the 2004 Economic Census which records the educational composition of employment at the firm-level for five categories of education level. Firms are classified at the 4-digit CIC industry level. We aggregate the data to IO sector level and define $Share_{io}$ as the fraction of industry is employment with education level o. We measure the similarity of employments in industries i and j through the correlation of $Share_{io}$ and $Share_{jo}$ across education levels.

Following EGK, We also use occupational data from the 2003 General Social Survey of China to construct an alternative measure of $Share_{io}$ where the subscript o now stand for occupation. There are 168 different occupations in total. The drawback of this data is that it is only available at 2-digit CIC industry level which is coarser than 2-digit IO sector (although there is a perfect one-to-one or many-to-one concordance from 2-digit IO sector to 2-digit CIC industry). This is why we use it as a robustness check. Note that $Share_{io}$ is still measured at 2-digit IO sector level when using occupational data. We assume that IO sectors belonging to the same 2-digit CIC industry have the same occupational composition (therefor their relatedness in terms of occupational similarity is equal to 1).

3.4 Knowledge spillover

Finally, industries may locate near each other to benefit from knowledge spillover which may involve exchange of information or ideas between workers as well as business leaders (Marshall, 1920; Saxenian, 1996; Glaeser and Kahn, 2001; Arzaghi and Henderson, 2008). It is very difficult to measure to what extent industries exchange knowledge with each other. EGK use technology transfers or patent citations between industries to measure this spillover. However, as noted by themselves, these measures only capture the highest level of information flows, rather than worker level spillovers.

Due to the above concern and also due to data limitations (we do not have data on technology transfers or patent citations for Chinese industries), we employ an alternative measure of knowledge spillover between industries. We base our measure on the co-production pattern of multi-industry firms across industries. We conject that firms producing in different industries are more likely to integrate with each other if there are more information exchange between these industries. Since the market for information is quite incomplete, vertical or horizontal integration between firms can be used as a means to internalize these externalities (Dumais et al., 2010; Atalay et al., 2014). Specifically, we use the following equation to calculate the extent of integration between industries:

$$R_{ij} = \min(\frac{X^i_{ij}}{X_i}, \frac{X^j_{ij}}{X_j}) \tag{4}$$

where X^i_{ij} (X^j_{ij}) denotes the export of industry i (j) that is exported by multi-industry exporters who export both products of industry i and industry j. X_i (X_j) denotes the total export of industry i (j). We use exports instead of outputs here because outputs at firm-industry level are not available in the CIE data, while firm exports at 8-digit HS level are available from the Chinese Custom data. We use the firm-level exports data for year 2000 which is the earliest year when this data is available for us. We aggregate firm exports to 2-digit IO sector level and drop all export-import companies that do not engage in manufacturing but serve exclusively as intermediaries between domestic producers (buyers) and foreign buyers (suppliers).[11]

3.5 Other variables

One focus of this paper is to investigate whether industry relatedness affect industry with different factor intensities differently. We measure industry factor intensity in two ways: capital intensity and skill intensity. Capital intensity is measured as capital (fixed assets) labor ratio, while skill intensity is the share of employment with higher than senior high school education. Both capital and skill intensity are calculated using the 2004 Economic Census data. We assume that these two variables do not change over time. To control for endowment driven comparative advantage, we also include interaction terms between industry and region characteristics. We measure two dimensions of region characteristics: share of college students in total population and share of non-agricultural population.

Another focus of our research is whether the national level FDI encouragement policies have differential effects on RCA dynamics across regions due to the different industrial structure of regions. We also employ this excercise as our identification strategy to address the potential endogeneity

[11]Since the data do not directly flag trade intermediaries, we follow standard practice and use keywords in firm names to identify them (Ahn et al., 2011). We drop XXXX wholesalers that mediate XXXX of China's trade.

problem induced by reverse causality. Following Sheng and Yang (2013), we construct two dummies for FDI encouragement and FDI restriction at the 4-digit CIC industry level. Then we aggregate them to region-IO-sector level using as weights the employment share of each CIC industry within each IO-sector at the country level. We use the initial employment share in 1998 to avoid potential endogeneity problems.

4 Empirical results

We first replicate the regression in Hausmann and Klinger (2007):

$$RCA_{ikt+1} = \alpha + \beta_1 RCA_{ikt} + \beta_2 DensityRelate_{ikt} + X_{ikt} + \varphi_{it} + \eta_{kt} + \epsilon_{ikt} \tag{5}$$

where RCA_{ikt} is a dummy variable which equals to 1 if the RCA of industry i in region k in year t is larger than 1 and 0 otherwise. $DensityRelate_{ikt}$ is a density measure based on the overall relatedness (see Equation 3) between industry i and all other industries in region k in year t. The calculation of this density measure is based on Equation 1. X_{ikt} is a vector of interaction terms including: interaction terms between density measure and the factor intensity of industry i, and interaction terms between industry factor intensity and region endowments. The purpose of the first set of interaction terms is to see whether density affects industries with different factor intensity differently. The second set of interaction terms are to control for endowment-driven comparative advantage. φ_{it} and η_{kt} are industry-year and region-year fixed effects, respectively.

As argued before, the $DensityRelate_{ikt}$ measure does not distinguish between different sources of relatedness. Thus we decompose relatedness into four different sources and run the following regressions:

$$\begin{aligned} RCA_{ikt+1} = \alpha &+ \beta_1 RCA_{ikt} + \beta_2 DensityOutput_{ikt} + \beta_3 DensityInput_{ikt} \\ &+ \beta_4 DensityLabor_{ikt} + \beta_5 DensityKnowledge_{ikt} + X_{ikt} + \varphi_{it} + \eta_{kt} + \epsilon_{ikt} \end{aligned} \tag{6}$$

where $DensityOutput_{ikt}$, $DensityInput_{ikt}$, $DensityLabor_{ikt}$, and $DensityKnowledge_{ikt}$ are density measures based on output linkage, input linkage, labor force similarity and the extent of knowledge spillovers between industry i and all other industries in region k in year t. The calculation of these measures are also based on Equation 1.

In general, we expect that all the density measures should have positive effects on RCA since high density measures of a region for an industry imply that it is relatively easy for firms in that industry to find suppliers, customers, employees and benefit from knowledge spillovers from related industries in that region. While seeking for inputs suppliers and outputs consumers should be common motives for firms in all industries, we note that for $DensityLabor$ and $DensityKnowledge$, high values may also imply that competition in the regional labor market may be tough and the risk of technology leakage can be high (especially for firms in high-tech industries). Thus we do not expect that these density variables have equal effects on different industries.

One potential problem with the above regressions is that the evolution of the RCA of industry i in region k may also affect the dynamics of the industry structure of that region which in turn affect the density measure. Thus there is a potential revers causality issue. To address this problem we employ a Difference-in-Difference strategy by interacting the exogeneous nationwide FDI policy measures (as defined in Section 3.5) with regional density measures. In a separate set of regressions, we add these interaction terms to the above regressions.

We first define region at the province level and in subsequent analysis we also run the regressions at the city level.[12] It turns out that the two sets of results are highly consistent.

4.1 General relatedness and RCA

Table 1 reports the province-level estimation results for Equation 5. All the standard errors are clustered at industry-year level. The density measure has been demeaned and divided by sample standard deviation. Column 1 shows that the density measure is highly significantly related with future RCA of an industry, i.e., industries with higher relatedness with other existing RCA industries in the region are more likely to gain or maintain RCA in that region. The magnitude of the effect is also economically significant. One standard deviation increase in density measure is related with 15 percentage points increase in the probability of gaining/maintaining RCA in the next year. This finding is consistent with Hausmann and Klinger (2007).

In column 2 to 4 of Table 1, we add interaction terms between density measure and industry capital and skill intensity. The interaction terms are positive but not significant. We will see in the next Section that this result is because that the density measure we use here does not distinguish between different sources of relatedness. In column 5 and 6, we further control for interaction terms

[12]There are 31 provinces and 287 cities in our sample.

between industry factor intensity and provincial endowment measures. This does not change the main results and the interaction term between industry skill intensity and provincial skill endowment measures are significant and positive, as expected.

[Insert Table 1 here]

4.2 Sources of relatedness and RCA

Column 1 of Table 2 reports the province-level estimation results for Equation 6 where we decompose relatedness between industries into four different sources. All our four density measures are significantly positively correlated with RCA. The coefficients are also economically significant. Note that all the density variables have been demeaned and divided by their sample standard deviation. The results indicate that one standard deviation increase in the output or the input density will increase the probability of industry i to have RCA in region k in year $t+1$ by 1.3 percentage point, while the same change in labor-market-pooling density or knowledge-spillover density will increase the probability by 11.6 or 6.9 percentage points, respectively. These results reveal that the interaction between industrial relatedness and regional industrial structure is an important source of regional comparative advantage which has not been fully considered in traditional trade theories (Hidalgo et al., 2007; Hausmann et al., 2014). These results also echoes the finding of EGK that the coagglomeration patterns of industries across US regions are affected by the relatedness between industries.

We add interaction terms between the density measures and industry capital and skill intensity in column 2 to 4. The results show that industries with higher capital intensity or skill intensity benefit more from being located in the same province with other industries using similar labor (in terms of educational composition) and benefit less from knowledge spillover from other local industries. One possible explanation for these results is that more capital or skill intensive industries normally use better-quality workers and thus may benefit more from improved employer-employee matching resulting from a larger regional labor pool. Meanwhile, these industries may benefit less from knowledge spillover since they normally use better technology than other industries. Moreover, the expectation that other firms in related industries may learn from them and become their competitors in the future may deter new entrants of these industries. Nevertheless, we find that even for the most skill intensive industry (with skill intensity 0.32), the overall effect of knowledge spillover is still postive, although it is much smaller than that for labor intensive industries.[13]

[13]From the results in column 3 of Table 2, we can calculate the effect of knowledge spillover on the RCA of the most skill intensive industry to be: $0.1 - 0.32 \times 0.24 = 0.023$

Column 3 also show that more skill intensive industries benefit more from higher input density. This may due to that these industries may require more customized inputs which makes geographical proximity more useful. Finally, being close to buyers of their outputs seems to be equally important for all industries. Overall, these results show that industries with different factor intensity indeed are affected differently by their relatedness with other local industries.

In column 5 and 6 of Table 2, we further add interaction terms between industry factor intensities and regional endowment measures to control for endowment driven comparative advantage. We can see that the interaction terms between industry skill intensity and these two endowment measures are highly positively significant which confirms the relevance of H-O forces. But notably, our main results discussed above do not change much after controlling for these variables.

In Table 7 in the Appendix, we re-run all the regressions in Table 2 except that we now use correlation of occupational composition (instead of educational composition) between industries in calculating labor density. Overall, the results are similar. The results for labor density and its interaction terms are less significant. This may be caused by measurement error problems since occupation data is only available at 2-digit CIC industry level which is more aggregate than the I-O sector level so that I-O sectors within the same 2-digit CIC industry are assigned the same occupational composition.

[Insert Table 2 here]

Table 3 reports the results for FDI policies. We focus on interaction terms between FDI policy measures (both encouragement and restriction) and the density measures to see whether these policies have different effects in regions with different industrial structures. As FDI policies are industry-year specific, their level effects on RCA can not be estimated since we have controlled for industry-year fixed effects.

The results in column 1 show that FDI encouragement policies tend to be more successful in regions with higher labor density, but less successful in regions with higher knowledge spillover density. The effects of FDI restriction policies have similar pattern but the coefficients are not significant. One concern with these results is that FDI policies may be correlated with industry factor intensities. Thus we control for interaction terms between densities and industry factor intensities in column 2 and find that the results do not change. In column 3 and 4 we further control for the H-O forces, the results are similar (except that the interaction term between FDI restriction and knowledge spillover is now significant). In Table 8 in the Appendix, we use industrial occupational composition instead of educational composition to measure similarity of labor types between industries. The results are similar, although less significant.

These results echoes the above findings in Table 2. FDI firms are normally more capital and skill intensive than domestic firms even within the same industry. Thus they may care more about the characteristics of the regional labor pool and care less about knowledge spillover from other industries (or rather worry more about knowledge spillover to other industries) for similar reasons as we have discussed above.

These results indicate that government policies designed to promote regional industry upgrading are more likely to be successful if they take into account the current industrial structure of the region and target industries that are most complementary with its current comparative advantage industries.

[Insert Table 3 here]

4.3 City-level results

Table 4 to 6 report city-level regression results. Overall, the results are highly consistent with what we have found in province-level regressions.

Table 4 shows that the general density measure plays even more important role for future RCA at the city level. One standard deviation increase in density is positively and significantly associated with 29.7 percentage points increase in the probability of gaining/maintaining RCA in the next year. Interaction terms between density and industry factor intensity show no robust results across different specifications. We believe that this is also because that the general density measure hides different sources of relatedness. As we have found in Table 2, interaction term between industry factor intensity and labor density is positively significant, while for knowledge density, it is negatively significant. This explains why it is difficult to find consistent results for these interaction terms when using general relatedness measure.

Table 5 confirms the results that more skill intensive industries benefit more from a larger labor pool, but less from knowledge spill over. It further shows that more skill intensive (as well as more capital intensive) industries benefit more from being close to their input suppliers.

Table 6 focus on the interaction terms between FDI policy and the various density measures. The results (especially that in column 3 and 4) show that FDI encouragement policy tend to be more successful in regions with higher output, input and labor density, while less successful in regions with higher knowledge density. These results are consistent with what we found at the province level.

Table 9 and 10 in the Appendix use occupation data to measure labor density. The results are similar as we just found in Table 5 and 6.

[Insert Table 4 here]

[Insert Table 5 here]

[Insert Table 6 here]

5 Conclusion

This paper presents empirical evidence from disaggregated Chinese industrial data to uncover how the revealed comparative advantage (RCA) of an industry in a region and industrial upgrading are affected by the relatedness between industries. To identify different sources of relatedness between industries, we quantify relatedness by employing the density measure developed by Hidalgo et al. (2007) and further decompose the sources of relatedness into four mechanisms: input linkage, output linkage, labor market pooling, and knowledge spillovers. To examine how the relations between relatedness and industrial upgrading are industry-heterogeneous, we categorize industries by their capital and skill intensities. We also exploit the FDI policy variation across industries in China to more carefully identify the effects of different mechanisms of industry relatedness.

There are three main findings. First, all four sources of relatedness significantly increase the future probability of an industry to have RCA in that region, and the future RCA of an industry is largely driven by its relatedness with other local industries regarding labor pooling and knowledge spillover instead of input-output linkages. Second, industries with higher capital or skill intensity, and thus, industry upgrading, benefit more from being located in the same region with other industries using similar labor, while benefit less from knowledge spillover from other local industries. In other words, the effect of knowledge spillover for capital intensive industry is much smaller than that for labor intensive industries. Third, the aforementioned findings are further confirmed by the tests using FDI policies in China: FDI encouragement policies tend to be more successful in regions with higher labor pooling density, but less successful in regions with higher knowledge spillover density. Our results are robust to alternative measures of labor pooling mechanism, different aggregation levels for regions (both province and city), and controlling for endowment driven comparative advantage.

References

Ahn, JaeBin, Amit K. Khandelwal, and Shang-Jin Wei, "The role of intermediaries in facilitating trade," *Journal of International Economics*, May 2011, *84* (1), 73–85.

Amiti, Mary, "Location of vertically linked industries: Agglomeration versus comparative advantage," *European Economic Review*, 2005, *49*, 809832.

Arzaghi, Mohammad and J. Vernon Henderson, "Networking off Madison Avenue," *Review of Economic Studies*, 2008, *75* (4), 101138.

Atalay, Enghin, Ali Horta?su, and Chad Syverson, "Vertical integration and input flows," *American Economics Review*, 2014, *104* (4), 11201148.

Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo, "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?," *Journal of International Economics*, 2014, *92* (1), 111–123.

Baldwin and Venables, "Relocating the value chain: Offshoring and agglomeration in the global economy," 2011. working paper.

Bao, Chengchao, Zhao Chen, and Jianfeng Wu, "Chinese manufacturing on the move: Factor supply or market access?," *China Economic Review*, 2013, *26* (C), 170181.

Blonigen, Bruce and Alyson Ma, "Please pass the catch-up: The relative performance of Chinese and foreign firms in Chinese exports," in Robert C. Feenstra and Shang-Jin Wei, eds., *China's Growing Role in World Trade*, Oxford: University of Chicago Press, 2010, pp. 339–351.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 2012, *97* (2), 339–351.

Cabral, Lus, Zhu Wang, and Daniel Yi Xu, "Competitors, complementors, parents and places: Explaining regional agglomeration in the U.S. auto industry," 2013. NBER Working Paper 18973.

Dumais, Guy, Glenn Ellison, and Edward L. Glaeser, "Geographic concentration as a dynamic process," 2010. NBER Working Paper 6270.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr, "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economics Review*, 2010, *100*, 11951213.

Glaeser, Edward L. and Matthew E. Kahn, "Decentralized employment and the transformation of the American city," 2001. NBER working paper 8117.

Hausmann, Ricardo and Bailey Klinger, "The structure of the product space and the evolution of comparative advantage," CID Working Paper 146, Center for International Development, Harvard University 2007.

_ , Cesar A. Hidalgo, Daniel P. Stock, and Muhammed A. Yildirim, "Implied comparative advantage," CID Working Paper 276, Center for International Development, Harvard University 2014.

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi, "Two Perspectives on Preferences and Structural Transformation," *American Economic Review*, December 2013, *103* (7), 2752–89.

_ , _ , and _ , "Growth and Structural Transformation," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 2B, Amsterdam and NewYork: North Holland, 2014, chapter 6, pp. 855–941.

Hidalgo, C.A., B. Klinger, A.-L. Barabási, and R. Hausmann, "The product space conditions the development of nations," *Science*, 2007, *317*, 482–487.

Ju, Jiandong, Justin Yifu Lin, and Yong Wang, "Industrial Dynamics, Endowment Structure, and Economic Growth,," Working Paper 2013.

Krugman, P. and A. J. Venables, "Integration, specialization and adjustment," *European Economic Review*, 1996, *40*, 959–967.

Marshall, Alfred, *Principles of Economics*, London: MacMillan, 1920.

Midelfart-Knarvik, Karen H., Henry G. Overman, and Anthony Venables, "Comparative advantage and economic geography: Estimating the determinants of industrial location in the EU," 2001. working paper.

Poncet, Sandra and Felipe Starosta de Waldema, "Product relatedness and firm exports in China," *The World Bank Economic Review*, 2013.

Puga and Venables, "The spread of industry-spatial agglomeration in economic development," *Journal of the Japanese and International Economies*, 1996, *10*, 440464.

Saxenian, AnnaLee, *Regional advantage: Culture and competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press, 1996.

Sheng, Liugang and Dennis T Yang, "Expanding Export Variety: The Role of Institutional Reforms in Developing Countries," IZA Discussion Papers 7611, Institute for the Study of Labor (IZA) September 2013.

Tang, Heiwai and Yifan Zhang, "Quality Differentiation and Trade Intermediation," Discussion Papers Series, Department of Economics, Tufts University 0771, Department of Economics, Tufts University 2012.

Venables, A. J., "The international division of industries-clustering and comparative advantage in a multi-industry model," *The Scandinavian Journal of Economics*, 1999, *101* (4), 495–513.

A Tables

Table 1: Coagglomeration-Based Relatedness - Province-Level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.775***	0.775***	0.775***	0.775***	0.773***	0.774***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Density	0.150***	0.148***	0.148***	0.148***	0.157***	0.154***
	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
× Industry capital intensity		0.028		0.026	0.036	0.050
		(0.040)		(0.044)	(0.043)	(0.043)
× Industry skill intensity			0.016	0.004	-0.064	-0.057
			(0.037)	(0.041)	(0.046)	(0.050)
Industry capital intensity					-0.554	
× Province college share					(0.620)	
Industry skill intensity					3.160***	
× Province college share					(0.672)	
Industry capital intensity						-0.258
× Province non-agriculture share						(0.200)
Industry skill intensity						0.572***
× Province non-agriculture share						(0.199)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26939	26939	26939	26939	26939	26939
R^2	.759	.759	.759	.759	.759	.759

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 2: Marshallian Relatedness - Province-Level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.782***	0.782***	0.781***	0.783***	0.781***	0.782***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Output linkage	0.013***	0.015***	0.017***	0.019***	0.020***	0.018***
	(0.003)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
× Industry capital intensity		-0.012		0.017	0.006	-0.002
		(0.038)		(0.059)	(0.059)	(0.059)
× Industry skill intensity			-0.033	-0.054	-0.049	-0.031
			(0.050)	(0.079)	(0.079)	(0.079)
Input linkage	0.013***	0.008*	0.002	0.001	-0.003	-0.002
	(0.003)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
× Industry capital intensity		0.054		0.026	0.037	0.035
		(0.053)		(0.059)	(0.059)	(0.059)
× Industry skill intensity			0.083**	0.070	0.079*	0.073
			(0.040)	(0.047)	(0.047)	(0.047)
Educational correlation	0.116***	0.075**	0.044	0.041	0.014	0.052
	(0.034)	(0.034)	(0.037)	(0.036)	(0.037)	(0.036)
× Industry capital intensity		0.222***		0.161***	0.164***	0.175***
		(0.052)		(0.058)	(0.057)	(0.057)
× Industry skill intensity			0.231***	0.156**	0.122*	0.109*
			(0.057)	(0.062)	(0.063)	(0.064)
Knowledge spillover	0.069***	0.091***	0.099***	0.103***	0.115***	0.119***
	(0.006)	(0.009)	(0.011)	(0.011)	(0.011)	(0.012)
× Industry capital intensity		-0.292***		-0.242*	-0.242*	-0.210
		(0.112)		(0.131)	(0.130)	(0.133)
× Industry skill intensity			-0.238***	-0.127	-0.206**	-0.255***
			(0.078)	(0.091)	(0.090)	(0.090)
Industry capital intensity					-0.379	
× Province college share					(0.689)	
Industry skill intensity					3.229***	
× Province college share					(0.683)	
Industry capital intensity						-0.253
× Province non-agriculture share						(0.227)
Industry skill intensity						0.899***
× Province non-agriculture share						(0.218)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26939	26939	26939	26939	26939	26939
R^2	.756	.756	.756	.756	.756	.756

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 3: FDI Policy and Marshallian Relatedness - Province-Level Regressions

	F.RCA			
VARIABLES	(1)	(2)	(3)	(4)
RCA	0.779***	0.779***	0.778***	0.778***
	(0.007)	(0.007)	(0.007)	(0.008)
FDI encouragement				
× Output linkage	0.002	0.003	0.002	0.003
	(0.006)	(0.006)	(0.006)	(0.006)
× Input linkage	0.011	0.009	0.009	0.009
	(0.007)	(0.007)	(0.007)	(0.007)
× Educational correlation	0.031***	0.029***	0.031***	0.031***
	(0.008)	(0.008)	(0.008)	(0.008)
× Knowledge spillover	-0.028**	-0.022*	-0.023**	-0.023*
	(0.012)	(0.012)	(0.012)	(0.012)
FDI restriction				
× Output linkage	0.085	0.086	0.111	0.130
	(0.105)	(0.108)	(0.109)	(0.109)
× Input linkage	0.068	0.076	0.079	0.082
	(0.056)	(0.056)	(0.056)	(0.056)
× Educational correlation	0.011	-0.009	-0.012	-0.007
	(0.059)	(0.059)	(0.059)	(0.058)
× Knowledge spillover	-0.068	-0.073	-0.087*	-0.093*
	(0.049)	(0.049)	(0.048)	(0.048)
Output linkage	0.013***	0.018**	0.019***	0.018**
	(0.004)	(0.007)	(0.007)	(0.007)
× Industry capital intensity		0.014	0.004	-0.004
		(0.061)	(0.061)	(0.060)
× Industry skill intensity		-0.052	-0.049	-0.032
		(0.082)	(0.081)	(0.081)
Input linkage	0.006	-0.001	-0.005	-0.004
	(0.005)	(0.006)	(0.006)	(0.006)
× Industry capital intensity		0.026	0.039	0.037
		(0.060)	(0.060)	(0.060)
× Industry skill intensity		0.047	0.056	0.049
		(0.048)	(0.048)	(0.048)
Educational correlation	0.085**	0.016	-0.015	0.026
	(0.035)	(0.037)	(0.037)	(0.036)
× Industry capital intensity		0.168***	0.170***	0.183***
		(0.059)	(0.059)	(0.058)
× Industry skill intensity		0.145**	0.110*	0.094
		(0.064)	(0.065)	(0.066)
Knowledge spillover	0.085***	0.115***	0.128***	0.132***
	(0.008)	(0.011)	(0.012)	(0.012)
× Industry capital intensity		-0.250*	-0.251*	-0.218
		(0.134)	(0.133)	(0.136)
× Industry skill intensity		-0.099	-0.180**	-0.234**
		(0.092)	(0.091)	(0.091)
Industry capital intensity			-0.365	
× Province college share			(0.689)	
Industry skill intensity			3.374***	
× Province college share			(0.678)	
Industry capital intensity				-0.263
× Province non-agriculture share				(0.229)
Industry skill intensity				0.973***
× Province non-agriculture share				(0.219)
Industry-year FEs	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes
Observations	26939	26939	26939	26939
R^2	.756	.756	.756	.756

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 4: Coagglomeration-Based Relatedness - City-level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.789***	0.789***	0.789***	0.789***	0.769***	0.767***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Density	0.297***	0.299***	0.299***	0.306***	0.580***	0.580***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.018)	(0.018)
× Industry capital intensity		0.028**		0.043***	-0.018	-0.034
		(0.013)		(0.016)	(0.022)	(0.021)
× Industry skill intensity			-0.013	-0.035**	0.075***	0.142***
			(0.014)	(0.017)	(0.023)	(0.022)
Industry capital intensity					-0.000*	
× Province college share					(0.000)	
Industry skill intensity					0.001***	
× Province college share					(0.000)	
Industry capital intensity						0.071
× Province non-agriculture share						(0.054)
Industry skill intensity						0.242***
× Province non-agriculture share						(0.059)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249403	249403	249403	249403	215512	223333
R^2	.715	.715	.715	.715	.698	.698

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 5: Marshallian Relatedness - City-Level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.791***	0.790***	0.791***	0.791***	0.760***	0.759***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Output linkage	0.011***	0.012***	0.016***	0.017***	0.015***	0.015***
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		-0.013		0.028	0.010	0.008
		(0.013)		(0.021)	(0.025)	(0.025)
× Industry skill intensity			-0.038**	-0.080***	-0.032	-0.031
			(0.019)	(0.031)	(0.036)	(0.035)
Input linkage	0.011***	0.008***	0.004	0.003	0.001	0.002
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.046*		0.018	0.058**	0.049*
		(0.024)		(0.025)	(0.025)	(0.026)
× Industry skill intensity			0.056***	0.058***	0.061***	0.062***
			(0.018)	(0.018)	(0.020)	(0.020)
Educational correlation	0.098***	0.098***	0.097***	0.103***	0.074***	0.087***
	(0.019)	(0.018)	(0.021)	(0.020)	(0.022)	(0.022)
× Industry capital intensity		0.047*		0.008	-0.025	-0.044
		(0.025)		(0.027)	(0.038)	(0.038)
× Industry skill intensity			0.107***	0.100***	0.342***	0.394***
			(0.028)	(0.031)	(0.042)	(0.042)
Knowledge spillover	0.005**	0.012***	0.027***	0.027***	0.084***	0.081***
	(0.002)	(0.004)	(0.005)	(0.005)	(0.007)	(0.007)
× Industry capital intensity		-0.063		0.018	0.103	0.107
		(0.045)		(0.051)	(0.071)	(0.071)
× Industry skill intensity			-0.171***	-0.161***	-0.262***	-0.247***
			(0.033)	(0.037)	(0.054)	(0.053)
Industry capital intensity					-0.000***	
× Province college share					(0.000)	
Industry skill intensity					0.001***	
× Province college share					(0.000)	
Industry capital intensity						0.014
× Province non-agriculture share						(0.057)
Industry skill intensity						0.246***
× Province non-agriculture share						(0.060)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249403	249403	249403	249403	215512	223333
R^2	.715	.715	.715	.715	.697	.697

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 6: FDI Policy and Marshallian Relatedness - City-Level Regressions

	F.RCA			
VARIABLES	(1)	(2)	(3)	(4)
RCA	0.790***	0.790***	0.759***	0.758***
	(0.004)	(0.004)	(0.004)	(0.004)
FDI encouragement				
× Output linkage	-0.003	-0.004*	0.004*	0.005**
	(0.002)	(0.002)	(0.003)	(0.003)
× Input linkage	0.007***	0.006**	0.008***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)
× Educational correlation	0.005	0.005	0.022***	0.023***
	(0.005)	(0.005)	(0.006)	(0.006)
× Knowledge spillover	-0.007	-0.004	-0.029***	-0.030***
	(0.005)	(0.005)	(0.007)	(0.007)
FDI restriction				
× Output linkage	0.050	0.072	0.053	0.016
	(0.054)	(0.055)	(0.056)	(0.061)
× Input linkage	-0.001	0.007	0.032	0.031
	(0.018)	(0.018)	(0.020)	(0.020)
× Educational correlation	0.026	0.026	-0.032	-0.031
	(0.036)	(0.035)	(0.048)	(0.047)
× Knowledge spillover	-0.012	-0.026	-0.011	0.004
	(0.029)	(0.028)	(0.041)	(0.041)
Output linkage	0.013***	0.019***	0.014***	0.014***
	(0.002)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.032	0.010	0.003
		(0.022)	(0.026)	(0.026)
× Industry skill intensity		-0.084***	-0.040	-0.033
		(0.032)	(0.037)	(0.036)
Input linkage	0.007***	0.001	-0.002	-0.001
	(0.002)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.014	0.057**	0.048*
		(0.025)	(0.025)	(0.026)
× Industry skill intensity		0.049***	0.044**	0.046**
		(0.019)	(0.021)	(0.020)
Educational correlation	0.102***	0.108***	0.056**	0.070***
	(0.020)	(0.020)	(0.022)	(0.022)
× Industry capital intensity		0.018	-0.021	-0.041
		(0.028)	(0.039)	(0.039)
× Industry skill intensity		0.094***	0.337***	0.389***
		(0.031)	(0.044)	(0.043)
Knowledge spillover	0.009**	0.029***	0.094***	0.091***
	(0.003)	(0.006)	(0.007)	(0.007)
× Industry capital intensity		0.009	0.113	0.119*
		(0.053)	(0.073)	(0.072)
× Industry skill intensity		-0.151***	-0.234***	-0.220***
		(0.038)	(0.054)	(0.053)
Industry capital intensity			-0.000***	
× Province college share			(0.000)	
Industry skill intensity			0.001***	
× Province college share			(0.000)	
Industry capital intensity				0.014
× Province non-agriculture share				(0.057)
Industry skill intensity				0.253***
× Province non-agriculture share				(0.060)
Industry-year FEs	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes
Observations	249403	249403	215512	223333
R^2	.715	.715	.697	.697

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Appendix

Table 7: Marshallian Relatedness (Occupation Data) - Province-Level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.780***	0.783***	0.782***	0.784***	0.783***	0.782***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Output linkage	0.012***	0.013***	0.015**	0.015**	0.017***	0.016**
	(0.003)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
× Industry capital intensity		-0.007		0.012	0.002	-0.010
		(0.038)		(0.059)	(0.059)	(0.059)
× Industry skill intensity			-0.023	-0.038	-0.037	-0.018
			(0.050)	(0.079)	(0.079)	(0.079)
Input linkage	0.008***	0.003	0.001	0.000	-0.003	-0.003
	(0.003)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
× Industry capital intensity		0.059		0.036	0.043	0.045
		(0.055)		(0.061)	(0.061)	(0.061)
× Industry skill intensity			0.058	0.042	0.058	0.052
			(0.042)	(0.048)	(0.047)	(0.047)
Occupational correlation	0.031***	0.018**	0.010	0.005	0.008	0.006
	(0.006)	(0.007)	(0.010)	(0.011)	(0.010)	(0.011)
× Industry capital intensity		0.174**		0.120	0.134*	0.142**
		(0.069)		(0.073)	(0.072)	(0.072)
× Industry skill intensity			0.160***	0.125*	0.090	0.098
			(0.062)	(0.065)	(0.066)	(0.066)
Knowledge spillover	0.066***	0.082***	0.093***	0.095***	0.108***	0.115***
	(0.006)	(0.008)	(0.010)	(0.010)	(0.010)	(0.011)
× Industry capital intensity		-0.275**		-0.185	-0.207	-0.188
		(0.115)		(0.133)	(0.130)	(0.134)
× Industry skill intensity			-0.245***	-0.166*	-0.238***	-0.290***
			(0.073)	(0.085)	(0.085)	(0.086)
Industry capital intensity					0.164	
× Province college share					(0.688)	
Industry skill intensity					3.241***	
× Province college share					(0.670)	
Industry capital intensity						-0.047
× Province non-agriculture share						(0.230)
Industry skill intensity						0.819***
× Province non-agriculture share						(0.215)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26939	26939	26939	26939	26939	26939
R^2	.756	.756	.756	.756	.756	.756

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 8: FDI Policy and Marshallian Relatedness (Occupation Data) - Province-Level Regressions

	F.RCA			
VARIABLES	(1)	(2)	(3)	(4)
RCA	0.779***	0.783***	0.781***	0.781***
	(0.007)	(0.007)	(0.007)	(0.007)
FDI encouragement				
× Output linkage	0.001	0.002	0.002	0.001
	(0.006)	(0.007)	(0.007)	(0.007)
× Input linkage	0.008	0.006	0.006	0.006
	(0.007)	(0.008)	(0.008)	(0.008)
× Occupational correlation	0.021**	0.018*	0.018*	0.018*
	(0.010)	(0.010)	(0.010)	(0.010)
× Knowledge spillover	-0.024**	-0.016	-0.017	-0.016
	(0.011)	(0.012)	(0.012)	(0.012)
FDI restriction				
× Output linkage	0.143	0.151	0.167	0.190*
	(0.100)	(0.104)	(0.104)	(0.105)
× Input linkage	0.043	0.047	0.046	0.050
	(0.059)	(0.059)	(0.059)	(0.059)
× Occupational correlation	0.073	0.070	0.077	0.081
	(0.084)	(0.083)	(0.082)	(0.082)
× Knowledge spillover	-0.118**	-0.122**	-0.138**	-0.147**
	(0.059)	(0.058)	(0.057)	(0.058)
Output linkage	0.011**	0.015**	0.017**	0.016**
	(0.005)	(0.007)	(0.007)	(0.007)
× Industry capital intensity		0.017	0.008	-0.004
		(0.061)	(0.061)	(0.061)
× Industry skill intensity		-0.050	-0.051	-0.033
		(0.082)	(0.081)	(0.081)
Input linkage	0.004	-0.001	-0.004	-0.004
	(0.005)	(0.007)	(0.007)	(0.007)
× Industry capital intensity		0.038	0.046	0.047
		(0.061)	(0.061)	(0.062)
× Industry skill intensity		0.022	0.038	0.031
		(0.050)	(0.049)	(0.048)
Occupational correlation	0.019**	-0.004	-0.001	-0.004
	(0.009)	(0.012)	(0.012)	(0.012)
× Industry capital intensity		0.124*	0.139*	0.148**
		(0.074)	(0.072)	(0.073)
× Industry skill intensity		0.117*	0.079	0.086
		(0.065)	(0.066)	(0.067)
Knowledge spillover	0.078***	0.101***	0.114***	0.122***
	(0.007)	(0.010)	(0.010)	(0.011)
× Industry capital intensity		-0.195	-0.218	-0.196
		(0.137)	(0.133)	(0.138)
× Industry skill intensity		-0.130	-0.202**	-0.262***
		(0.087)	(0.087)	(0.088)
Industry capital intensity			0.115	
× Province college share			(0.689)	
Industry skill intensity			3.327***	
× Province college share			(0.665)	
Industry capital intensity				-0.084
× Province non-agriculture share				(0.231)
Industry skill intensity				0.885***
× Province non-agriculture share				(0.215)
Industry-year FEs	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes
Observations	26939	26939	26939	26939
R^2	.756	.756	.756	.756

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 9: Marshallian Relatedness (Occupation Data) - City-Level Regressions

	F.RCA					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RCA	0.789***	0.788***	0.790***	0.789***	0.760***	0.760***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Output linkage	0.011***	0.012***	0.015***	0.016***	0.014***	0.015***
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		-0.017		0.023	0.013	0.012
		(0.014)		(0.022)	(0.025)	(0.025)
× Industry skill intensity			-0.033*	-0.075**	-0.043	-0.045
			(0.019)	(0.031)	(0.036)	(0.035)
Input linkage	0.009***	0.004**	0.002	0.000	0.001	0.003
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.059***		0.034	0.067***	0.059**
		(0.023)		(0.024)	(0.024)	(0.025)
× Industry skill intensity			0.057***	0.052***	0.025	0.018
			(0.018)	(0.019)	(0.020)	(0.020)
Occupational correlation	0.016***	0.015***	0.019***	0.021***	0.006	0.001
	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
× Industry capital intensity		0.024		0.004	-0.003	-0.010
		(0.038)		(0.040)	(0.040)	(0.040)
× Industry skill intensity			-0.002	-0.002	0.198***	0.246***
			(0.028)	(0.030)	(0.039)	(0.038)
Knowledge spillover	0.006***	0.010***	0.016***	0.017***	0.071***	0.066***
	(0.002)	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)
× Industry capital intensity		-0.039		0.043	0.104	0.109
		(0.046)		(0.052)	(0.072)	(0.071)
× Industry skill intensity			-0.097***	-0.101***	-0.205***	-0.173***
			(0.029)	(0.033)	(0.047)	(0.046)
Industry capital intensity					-0.000	
× Province college share					(0.000)	
Industry skill intensity					0.001***	
× Province college share					(0.000)	
Industry capital intensity						0.046
× Province non-agriculture share						(0.056)
Industry skill intensity						0.244***
× Province non-agriculture share						(0.059)
Industry-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249403	249403	249403	249403	215512	223333
R^2	.715	.715	.715	.715	.697	.697

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01

Table 10: FDI Policy and Marshallian Relatedness (Occupation Data) - City-Level Regressions

	F.RCA			
VARIABLES	(1)	(2)	(3)	(4)
RCA	0.788***	0.788***	0.760***	0.759***
	(0.004)	(0.004)	(0.004)	(0.004)
FDI encouragement				
× Output linkage	-0.004	-0.004*	0.004	0.005*
	(0.002)	(0.002)	(0.003)	(0.003)
× Input linkage	0.004	0.003	0.005*	0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
× Occupational correlation	0.013***	0.011**	0.020***	0.022***
	(0.004)	(0.004)	(0.005)	(0.005)
× Knowledge spillover	-0.014***	-0.012**	-0.030***	-0.032***
	(0.005)	(0.005)	(0.007)	(0.007)
FDI restriction				
× Output linkage	0.048	0.087	0.097*	0.060
	(0.056)	(0.058)	(0.058)	(0.062)
× Input linkage	0.038**	0.044**	0.034*	0.035*
	(0.018)	(0.018)	(0.019)	(0.019)
× Occupational correlation	-0.043*	-0.032	-0.014	-0.021
	(0.024)	(0.025)	(0.045)	(0.042)
× Knowledge spillover	0.012	-0.007	-0.026	-0.006
	(0.025)	(0.028)	(0.036)	(0.035)
Output linkage	0.013***	0.019***	0.014***	0.015***
	(0.002)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.029	0.016	0.010
		(0.023)	(0.026)	(0.026)
× Industry skill intensity		-0.088***	-0.057	-0.053
		(0.032)	(0.037)	(0.036)
Input linkage	0.007***	0.000	0.000	0.002
	(0.002)	(0.003)	(0.003)	(0.003)
× Industry capital intensity		0.034	0.067***	0.059**
		(0.024)	(0.024)	(0.025)
× Industry skill intensity		0.038*	0.007	0.001
		(0.020)	(0.021)	(0.021)
Occupational correlation	0.011***	0.016***	-0.004	-0.009
	(0.003)	(0.005)	(0.007)	(0.007)
× Industry capital intensity		0.001	-0.007	-0.017
		(0.042)	(0.041)	(0.041)
× Industry skill intensity		0.002	0.195***	0.242***
		(0.028)	(0.039)	(0.038)
Knowledge spillover	0.012***	0.019***	0.079***	0.074***
	(0.003)	(0.005)	(0.006)	(0.006)
× Industry capital intensity		0.051	0.123*	0.131*
		(0.054)	(0.073)	(0.072)
× Industry skill intensity		-0.083**	-0.172***	-0.139***
		(0.034)	(0.047)	(0.046)
Industry capital intensity			-0.000*	
× Province college share			(0.000)	
Industry skill intensity			0.001***	
× Province college share			(0.000)	
Industry capital intensity				0.040
× Province non-agriculture share				(0.056)
Industry skill intensity				0.251***
× Province non-agriculture share				(0.059)
Industry-year FEs	Yes	Yes	Yes	Yes
Province-year FEs	Yes	Yes	Yes	Yes
Observations	249403	249403	215512	223333
R^2	.715	.715	.697	.697

Robust standard errors clustered at industry-year level in parentheses
* p< 0.1, ** p< 0.05, *** p< 0.01