Agglomeration, Public Expenditure and Productivity Spillover in Space: Firm-Level Evidence from China's Electric Apparatus Industry

Paper presented at the 2016 AEA Annual Meeting

Ming He, Yang Chen, and Ron Schramm*
IBSS, Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu Province 215123, China

January 5, 2016, San Francisco

Abstract

The article consists of two parts. Part I examines the *Chinese Annual Survey of Industrial Firms database* (CASIF). We start with a brief description of the coverage and key variables, followed by a review of recent applications of the data set in empirical studies. We then discuss the quality and the limitations of the dataset and propose a data cleaning procedure that identifies the key information of each firm. In Part II, linking a sub-sample of this micro-level data set with two other databases-Chinese Prefecture City and County Fiscal Statistical Material and county geographic information, we explore the productivity propagation process in space in China's electric apparatus industry during 1999-2007. The impacts of firm-specific characteristics on productivity growth, the external market conditions and the institutional factor are all considered. Within the general nested spatial framework, we propose modifications to the Kelejian and Prucha's (1998) FE-2SLS procedure and the Mutl and Pfaffermayr's (2011) RE-FG2SLS procedure to cope with the unbalanced panel and find the statistical evidence strongly favors the fixed effects over the random effects model. The main results indicate (1) there are significantly positive within-region as well as between-region productivity spillovers; (2) market competition and public expenditure in the local and neighboring districts/counties are important sources to boost productivity growth; and (3) the externality in productivity growth attenuates in spatial distance. Notably, the between-regional spillover effects are found to be more significant in smaller than in larger administrative units and more pronounced in urban districts than in counties.

*Ming He: ming.he@xjtlu.edu.cn, +86-0512-8816-1712; Yang Chen: yang.chen@xjtlu.edu.cn, +86-0512-8816-1178; Ron Schramm: ron.schramm@xjtlu.edu.cn, +86-0512-8816-7129.

Part I
Working with the Chinese Annual Survey of Industrial Firms Database

1 Basic information about CASIF

The *Chinese Annual Survey of Industrial Firms database* (CASIF) is conducted by National Bureau of Statistics (NBS), similar to the Longitudinal Research Database (LRD) maintained by the U.S. Bureau of the Census and the Annual Census of Production Respondents Database (ARD) and Annual Enterprise Survey (AES) in New Zealand. This data set covers all state-owned enterprises (SOEs) and non-state-owned firms with annual turnover over 5 million RMB (also referred to as above-scale industrial firms). [1]

According to the industrial classification specified by GB/T 4754-2002, CASIF covers industrial firms in the mining sector (0610-1100), manufacturing sector (1310-4320), and the public utilities sector (4411-4620). The names of two-digit sectors are shown in Table 1. Each observation in the data set is a firm, termed as a "legal unit."[2] Note that the entities defined in CASIF differs from the concept of plants used by other countries' firm-level surveys, where each entity is a physical establishment operating at a particular address.[3]

Our version of the CASIF starts from 1996 and ends in 2010, with a total of 3,052,464 observations on 615,624 distinct firms. Annually, the number of observations ranges from 159,703 to 334,151 during the period of 1998-2008. In Table 2 we report the number of firms, total output, and total employment for each year in the survey period. We also compute these values for all above-scale firms in the data set. These numbers are then compared to those reported by the *China Statistical Yearbook* (2012 issue). It can be seen that the share of the small firms in the full sample is about 10% during the period of 1998-2009. They contribute a very small fraction (roughly 1%) of total output. In 2002, 2003, 2005, 2006, and 2007, the total

[1]Note that (1) the inclusion threshold is not strictly enforced. Many firms sampled in one year with sales lower than the threshold in the following year are not required to report to the annual survey but could still continue their reporting (5% of private or collectively-owned firms compared to 19% of SOEs); (2) the sample section of this firm-level dataset at the lower threshold is biased toward highly productive small firms given their low employment level; and (3) the threshold for non-SOEs was increased to 20 million RMB in 2011.

[2]The individual firms in the sample could be the subsidiaries affiliated to large enterprises if (1) they are established legally, having their names, organizations, geographical location and capability to take civil liability; (2) possessing and using assets independently, assuming liabilities and are entitled to sign contracts with others; and (3) accounting independent and compile their own balance sheets. Other "industrial activity units" which cannot take civil liability are not included.

[3]For example, the official U.K. Annual Census of Production Respondents collects plant-level information reported by each firm.

output and employment of above-scale firms computed from the CASIF are almost identical to those in the statistical yearbook, the differences being less than 5%. The coverage of CASIF over these years is fairly good. The discrepancies between CASIF and the yearbook are larger at both ends of the sample period. For instance, 54% of all above-scale firms are missing from CASIF in 2009, and the other aggregates are also lower than the national values. In 2004, the discrepancy is small for total output but relatively high in the number of firms and total employment. In sum, the coverage of CASIF seems to be good between 1998 and 2007.[4]

Most researchers use the sample from 1998 to 2007 also because the annual data sets in other years have compatibility issues. The sample size is very small prior to 1998. The 2008-2010 data misses important financial variables such as fixed assets, intermediate inputs, wages, revenue, etc. The quality of the data is also questionable after 2008. Judged by the number of firms, total output, and total employment, there has been a sharp decline in the coverage in the 2008 and 2009 data. Oddly, the number of new entrants in 2009 is zero. The total employment and output of above-scale firms from the 2010 data far exceed the national aggregates reported by the *China Statistical Yearbook*.

The CASIF shows large scale entry/exit behavior on an annual basis. During the 1998-2007 period, the number of new entrants as a fraction of the annual sample, ranges from 14.4% to 45.6% (Table 2). Similarly, the percent of exits ranges from 8.2% to 25.8%. Out of the 577,649 distinct firms that are present in the 1998-2007 sample, 339,407 have continuous observations for three or more years. Only 164,529 firms survive five years or more. These numbers suggest that the CASIF data is highly unbalanced in its panel representation.

The CASIF provides identifying information and accounting/financial information for each observation. The key variables are shown in Table 3. Here we make a few remarks. First, some financial variables such as R&D expenses are valuable for specific research topics, but they are not consistently reported in the dataset. R&D expenditures are available for six year from 2001-2007 (not including 2004) and the number of computers (not reported) in 2004 only. Various employment benefits beyond salaries including labor and unemployment insurance payable, pension and medical insurance payable, as well as housing provision fund and subsidy are not reported prior to 2004. Similarly, miscellaneous expenditure categories including advertising expenses, transportation expenses, and employee training expenses are not reported before 2004. Oddly, the 2004 annual data set does not provide a few critical financial variables such as gross output, output of new merchandise, total sales revenue, export sales, etc. Decomposition of the labor force by education levels (postgraduate, university, college, high school, primary or less), technical titles, the net cash flow, parent firm name, etc., are only available in 2004. The cash flow variables are not available in 1998-2003, but both inflow and outflow volumes are reported in 2005-2009. A few variables are coded, including operation status, and

[4]Census data of all industrial firms regardless of ownership and size is available (but not publicly accessible) for years 2004, 2008 and 2013. Brandt, Van Biesebroeck, and Zhang (2014) compared the 2004 and 2008 census data with CASIF.

affiliation type. The codings are explained in Tables 4 and 5.

The other two popular sources of Chinese firm-level data are Wind and CSMAR (China Stock Market and Accounting Research).[5] Both databases provide complete financial data (annual/interim/quarterly reports) of all listed companies in the Shanghai Stock Exchange and the Shenzhen Stock Exchange. Historical data traces back to 1990. The CSMAR database compiles variables from four financial statements: the balance sheet, the income statement, the statement of cash flow (direct/indirect method). The number of observations ranges from 926 in 1998 to 2736 in 2014. The Wind database also provides various identification information such as the stock code, the business name, the industry code, and the zip code. In comparison with CASIF, these two databases provide extra accounting and financial variables at higher frequency and with a wider coverage of industrial sectors.

2 How CASIF is used by economists

The CASIF has been used in a number of fields, including (but not restricted to) international trade, industrial organization, labor and macroeconomics. This section reviews some typical applications.

The data set has been widely used to estimate firm-level productivity. The traditional approach is to estimate the Solow residual by regressing a Cobb-Douglas production function. The least squares estimator, however, is known to suffer from endogeneity bias and sample selection bias because firms may adjust the variable inputs in accordance with innovations in technology and may even exit the market when the operation is no longer profitable. These problems are solved by Olley and Pakes's (1996) and Levinsohn and Petrin's (2003) semi-parametric estimators. Crucial to these methods is information on investment (Olley and Pakes, 1996), intermediate inputs (Levinsohn and Petrin, 2003), and exit behavior (Olley and Pakes, 1996). In practice, both methods have been used. An often overlooked issue is the measure of output and input. Ideally, these quantities should be measure in real terms. Except for the gross output and employment, the CASIF reports the value of all other inputs at current prices, but a price index is missing. Brandt, Van Biesebroeck, and Zhang (2012) constructed two price indices. Their output deflator is based on the price index from the *China Statistical Yearbook*, while the input deflators are computed from the price indices of various industrial inputs and China's national Input-Output Table. Labor productivity (Jefferson, G., Li, and Zheng, 2000; Yueh, 2010) is the next frequently used measure of productivity if the research is focused on wage and employment. However, this indicator ignores the crucial role of capital (De Loecker, 2007) in determining a firm's technical status and productive efficiency.

This data set is also widely used to identify the effects of FDI. Overall, Abraham, Konings, and Slootmaekers (2010) find positive spillovers by FDI in the Chinese manufacturing industry, but the spillovers from Hongkong-Maucau-Taiwan (HMT) invested firms are found to be negative. In a similar work, Lin, Liu, and Zhang

[5]Wind is provided by Wind Info Co., Ltd and CSMAR is jointly provided by GTA Information Technology Co. Ltd, the University of Hong Kong and the Hong Kong Polytechnic University.

(2009) demonstrate negative horizontal spillover effects from HMT firms and positive horizontal spillover effects of FDI from most OECD countries. The impact on local firm productivity not only varies with the origin of the FDI but also depends on the time frame. For example, using CASIF, Liu (2008) finds out that an increase in FDI in four-digit industrial sectors improves the long-term productivity growth of domestic firms but lowers their short-term productivity. Instead of focusing on horizontal or vertical effects within an industrial sector, Jeon, Park, and Ghauri (2013) test the heterogeneous effects of FDI on local firm's productivity across industries, finding that FDI are more likely to generate negative influences on local Chinese firms, especially in low-technology sectors while the effects of FDI on other industries are found to be positive. Following the release of outward direct investment (ODI) data by the Ministry of Commerce in recent years, researchers begin to probe into the relationship between productivity and ODI decisions. By merging the ODI data of the nation and that of Zhejiang province into CASIF, Tian and Yu (2014) estimate that a 1% increase in firm's productivity leads to a 4% increase in the size of ODI conditional on firm's engagement in ODI. A related strand of literature explores the relationship between productivity and trade. Lu (2010) shows that Chinese exporters are less productive, while Dai, Maitra, and Yu (2012) argue that this result is largely caused by the inclusion of China's processing exporters. They show that Chinese exporters are more productive than non-exporters if processing exporters are excluded. Researchers also study the sorting effect of trade on Chinese firms. Schmerer and Wang (2014) provide a recent study on this topic based on the CASIF.

There are a few studies linking firm performance with the ownership. For example, Jefferson, G., Li, and Zheng (2000) examine the differences in marginal factor productivity across foreign-linked, shareholding and private enterprises from 1980 to 1996, finding modest productivity outcome in non-state and non-collective and productivity declines in those shareholding enterprises. Similar results are found by Song and Yao (2004). They conclude that partial state control and private control both lead to higher profitability than state ownership, but there is little effect of these restructuring on unit cost and productivity. Jefferson and Su (2006) show the probability of ownership conversion increases with the firms productivity as well as the intensity of competition. These results indicate selection bias in the privatization process of Chinese SOEs, which is consistent with the government policy of releasing the smaller firms while retaining the larger ones. In addition to examining the direct linkage between the conversion of ownership and the change in firm productivity, a few studies explore the productivity-export, productivity-R&D association controlling the type of firm. Using a cross-sectional data set for Chinese enterprises for various ownership types, Hu (2001) finds insignificant contribution of government R&D expenditure to firm productivity compared with the investment from within the firm. Sun and Hong (2011), covering 70,000 Chinese firms during 2001-2005, find the foreign-owned exporters benefit less from exporting compared with the domestic traders. Du, Liu, and Zhou (2014) analyze the contribution of the state and non-state sectors in the aggregate total factor productivity to verify the recent debate on the existence and scale of Chinas state sector advancing and the private sector retreating. A recent study by Hu, Xu, and Yashiro (2015) explores, for the whole sample as well as for the sub-samples of firms with different ownerships (SOEs, private enterprises and foreign-invested enterprises), the benefit from agglomeration effects which acts as a source of externality for a wide range of industries and in 2860 counties. They argue that private enterprises are the main sources of agglomeration effects especially in upstream industries. With a dynamic point of view, firms with state capital, collective capital or corporate capital are reported to be more volatile in productivity in Luo and Zhu (2015).

China is also known for her phenomenal productivity and GDP growth as well as her poorly developed financial system. In an interesting study, Chen and Guariglia (2013) investigate the relationship between the liquidity constraint and firm-level productivity using a panel of 130,840 firms extracted from CASIF over the period of 2001-2007. Their work reveals a strong negative impact of the liquidity constraint on productivity, especially for foreign and private firms. CASIF has also been used to study various policy effects. Chandra and Long (2013) analyze the impact of the 2004 VAT tax rebate reform on Chinese manufacturing firm's exports. According to their estimates, exports increases by 13% following a 1% tax rebate. Gao and Van Biesebroeck (2014) find that the market reform of the electricity generation sector introduced in 2002 boosts labor and material efficiency.

3 Challenges and caveats

The CASIF is known (Nie, Jiang, and Yang, 2012) to contain numerous errors and internal inconsistencies. A major challenge we face is the identification problem: Even though the raw data provides detailed information on firm ID, geolocation, and industrial sector of each establishment, in many cases their values are non-unique in the survey period. This results from recording errors and revisions in the coding system. These key identifiers must be cleaned before we can build a longitudinal data set. Our data-cleaning procedure is described below.

3.1 Identification of unique firm ID

Each observation (establishment) is jointly identified by the organization code and the name of the business. The organization code is the official identification issued by the registration office, which should remain unchanged throughout the life cycle of the establishment. This unique organization code conceptually could be used to identify an establishment when the annual data sets are merged into a longitudinal one. However, there are two major problems with this variable. In many occasions, different identification codes are assigned to the same firm (verified by the name and other non-ID information) in different years; this might be attributed to the change in the boundary of the firm or its ownership structure, following either a restructuring or an acquisition. In addition, in the same annual data set, two or more establishments (usually from the same administrative unit) share the same organization codes. To resolve the indeterminacy in organization codes, we adopt the matching procedure proposed by Nie, Jiang, and Yang (2012) as follows.

1. We pool the observations from all annual data sets, and group them by organization code. The observations in each group share the same organization code. If observations from the same group have more than one business name, it indicates that the establishment changed its business name in the corresponding year.

2. For each group G constructed in the previous step, we find other groups that have at least one observation sharing the same business name with an observation from group G. Once these groups are identified, we append them to G, and remove duplicated observations from the resulting group G'.[6] The updated group G' may contain more than one organization code.

3. Repeat the previous step until the group structure no longer changes, then remove all duplicated groups. Each of the remaining group represents a unique establishment in the longitudinal data, to which a unique ID is assigned. We use the organization code associated with the group as the unique ID.

The raw data consists of 658,213 unique organization codes over 1996-2010. The algorithm converges in three iterations, retaining 620,020 groups. The numbers indicate that 6% of the full sample might have been misidentified as independent establishments without this procedure.[7]

Among the 620,020 unique establishments identified in the previous step, 8,825 have multiple observations in one or more years. About half of them (4,429) have more than two observations but only one multiplicity, in which case we compare the values of accounting and financial variables with those in adjacent years. The observation with a closer match is chosen and the other one is discarded. The observations associated with the remaining 4329 establishments are discarded.[8]

3.2 Identification of the geolocation

The geolocation of each establishment is identified by a six-digit administrative code. The data sets provide the six-digit (1996-2003) or twelve-digit (2004-2010) administrative code plus the six-digit zip code for each observation. The first six digits of the administration code define an administrative unit in the district/county level, while the last six digits define the township and community. In this study, we use the first six digits to identify the location of each establishment. We clean the data on administrative codes by the following procedure:

[6] Note that a group may be simultaneously appended to multiple groups, and that the group-combining operation may result duplicated groups that have identical set of observations.

[7] As far as the organization code and business name are changed one at a time, the iterated matching procedure is able to track down the same establishment over time, even if multiple changes take place in succession. However, the algorithm fails whenever an establishment changes the organization code and business name at the same time.

[8] 2,047 establishments have only two observations, both of which are recorded in the same year.

1. The administrative codes in the data are formatted with different versions of the GB/T 2260 standard.[9] We construct a junction table that maps earlier versions of the administrative code into the 2007 version. All valid administrative codes are then converted into the values specified by the 2007 version of the national standard (GB/T 2260-2007).

2. For the 16,771 observations that do not have valid administrative codes, we constructed a table that maps 37,249 zip codes to six-digit administrative codes (2007 version). Their administrative codes are then recovered from zip codes if the latter can be found in our mapping table.

3. If the previous steps result in a unique administrative code for an establishment in all years, it is used to identify the geolocation. 30,811 establishments are found to have multiple values, in which case we choose the most frequent one.

3.3 Identification of the industrial division

The industrial division of each establishment is identified by a four-digit industry code. The data set uses two different coding systems: GB/T 4754-1994 until 2002 and GB/T 4754-2002 afterwards. Old industry codes are mapped to their new values specified by GB/T 4754-2002. Similar to what we did with administrative codes, the industry codes are further cleaned so that all observations associated with an establishment are assigned a unique industry code.

Part II
The empirical study

4 Introduction

"Productivity isn't everything, but in the long run it is almost everything"–Paul Krugman

Measured as the efficiency in production to convert a set of inputs into the desired amount of output, and a key factor to explain the cross-regional differences in the level or growth of gross domestic product (GDP) (Hall and Jones, 1999; Easterly and Levine, 2002), the sources of total factor productivity (TFP) growth have been widely debated in the literature with both macro and micro perspective (see Syverson, 2010 for a survey and the references therein). The sources of productivity differences at

[9] The administrative divisions have undergone four major revisions in 1995, 1999, 2002, and 2007. Each time, a large number of administrative divisions were either renamed, merged, split, or (newly) created. In most cases, the administrative division affected was assigned a new six-digit code. The old administrative code was abolished and won't be used in the future. The 1995 version of the code map, GB/T 2260-1995 consists of 3404 distinct administrative codes, 1200 of which were revoked by 2007. The changes being made are tabulated in the appendix of the publication.

the micro level are usually attributed to varied elements such as the engagement in foreign markets, the firm's activity in innovation, the external market condition and the institutional environment. In most productivity analysis, researchers implicitly assume that the activity outside of a location has no effect on activity within the location. An early important departure from this tradition is provided by Rosenthal and Strange (2004) with a micro-level analysis of the geographic scope of agglomeration economies. Since then the producer practices that may have spillover effects within and across the geographic boundary on the productivity levels of others come into the lime light. These externalities are discussed in the context of classic agglomeration mechanisms such as input sharing, knowledge or technology spillover. Higher productivity correlations among "nearby" producers are usually tested by regressing the exogeneous variables such as R&D expenditure of other firms or the presence of foreign investment on the productivity level for a specific firm (Wei and Liu, 2006; Griffith, Redding, and Van Reenen, 2004; Keller and Yeaple, 2009). However, two research questions, although important, are not answered: (i) how much variation of the productivity growth could be attributed to the growth from a nearby neighbor; and (ii) whether the spillover effect generated from agglomeration through technology linkages or input sharing attenuates with geographic distance.

By answering these two questions, this paper contributes to the micro-level productivity analysis in several aspects. First, we apply a unique and extensive longitudinal dataset-CASIF described in Part I-combining fiscal data from Chinese Prefecture City and County Fiscal Statistical Material and county geographic information within the period 1999-2007. The assembled unbalanced longitudinal data set consists of 615,214 observations from 470 four-digit industrial sectors, located in 2,862 administrative unit (urban district or county). Second, the current literature on productivity analysis presumes that each sector has the same TFP, which is only true if the underlying production function has the same factor shares for different sectors. We instead estimate TFP for 84,727 observations (26,174 distinct firms) in the electronics apparatus sector (industrial code 3900) correcting for the sample selection and the simultaneity bias using the Levinsohn and Petrin (2003) framework and making the assumption that firms within the same sector face the same input price. Third, to measure the spatial interaction among firms in this sector, we define "neighborhood" using both the intra-regional and the inter-regional measure. Those in the same region (urban district or county) are defined as type-I neighbors and those located in the other region which either shares the same administrative boundary or within the 50km spherical distance are treated as the type-II neighbors. Fourth, to our knowledge, this is the first study to analyze the productivity spillover using the general nested spatial model Elhorst (2014) with inclusion of firm-specific heterogeneity in a huge unbalanced panel structure. Last but not the least, this paper finds out that public assets suggest significant within- and between-region productivity benefits to firms. And the contribution of public asset to productivity growth is positively associated with the size of the region where firms choose to locate.

Beyond adding the spatial dimension, we also look at the sources of productivity growth stemming from firm-specific attributes, including influences under the control of economic actors inside the business such as R&D and export. A firm with the more advanced technology would be more efficient than one with lagging production capabilities, leading to reduced costs and improved productivity. There is a long literature linking productivity with R&D activity (Griliches and Mairesse, 1991; Hu, 2001). Doraszelski and Jaumandreu (2013) model firm productivity growth as the consequence of R&D expenditures with uncertain outcomes using a panel of Spanish firms. Aw, Roberts, and Xu (2009) point out the bidirectional causality between R&D and productivity in their study of Taiwanese electronics exporters, showing that firms that select into exporting tend to be more productive but the decision to export is often accompanied by large R&D investments. Of course, R&D is just one of the more observable components to measure firm's overall innovative efforts. Some firms may undertake process and product innovation without formally reporting R&D spending. In comparison with previous studies using cross sectional data within a short time period (Hu and Jefferson, 2004; Hu, Jefferson, and Jinchang, 2005), Boeing, Mueller, and Sandner (2015) using a panel data of listed firms over two time periods 2001-2006 and 2007-2011, find out privately owned enterprises not only obtain higher returns from own R&D than majority and minority state-owned enterprises (SOEs), they are also able to increase their leading position. Compared to the existing literature, instead of using either the R&D expenditures or the share of R&D expenditure of gross output, we measure a firm's innovative activity using the share of new merchandise in the gross output, which contributes to 19.2 percent productivity growth in our base line fixed effects model.

The greater market-orientation (Jefferson and Su, 2006) as well as the "open door" policy with China joining the World Trade Organisation (WTO) in 2001 results in greater competitive pressures for firms to be engaged in foreign markets [10]. The empirical literature explaining firm performance and export behavior suggests two main mechanisms, namely, self-selection and learning by exporting. The self-selection mechanism is tested by Bao, Huang, and Wang (2015) for China, Clerides, Lach, and Tybout (1998) for Colombia, Mexico and Morocco, Bernard and Jensen (1999) for the US, Aw, Chung, and Roberts (2000) for Taiwan. The evidence in favor of the learning-enhanced productivity boosting hypothesis is documented in recent studies by Girma, Greenaway, and Kneller (2003) for UK and De Loecker (2007) for Slovenia. Wagner (2012) in his survey paper presents a large strand of literature on this topic but finds inconclusive evidence of the importance of exporting on productivity. For example, Silva, Afonso, and Africano (2012) reports in Portugal fast learning effects for exporters only to EU countries but no such effects for firms that export to less-developed nations; Pisu (2008) finds out no causal relation between exporting and productivity irrespective of development level of destination countries; Wilhelmsson and Kozlov (2007) find inconclusive evidence for learning-by-exporting as well. In the Chinese context, a productivity paradox has been reported by a few studies. Contrary to the self-selection hypothesis in the Melitz (2003) model, using CASIF during 1998-2005, Lu, Lu, and Tao (2010) find that among foreign affiliates, exporters are less

[10]This policy, on the one hand, attracts more foreign firms into the domestic market, increasing competition in the global economy and on the other hand, offers the opportunity for Chinese enterprises to obtain the technological know-how. The latter role played by openness builds the theoretical foundation of "catching up" analyzed in the economic growth models.

productive than non-exporters. A similar result is reported by Lu (2010), concluding that Chinese exporters are on average less productive than non-exporters. With these inconsistent results provided in the existing literature, to what extent exporting affects firm-specific TFP remains a relevant and interesting research question. Therefore, in this study, we control the firm's foreign market engagement using a measure of the fraction of gross sales exported. The impact of export on firm productivity is shown to be significantly positive in the electronics sectors in China, implying no such productivity paradox.

Measured as the share of a county's employment in a particulary industry, specialization does not significantly encourage productivity growth over the period 1956-1987 for the top six industries in a city (Glaeser, Kallal, Scheinkman, and Shleifer, 1992). We find similar results in this paper. The flip side of specialization is diversity, measured using a Herfindahl-Hirschman index (HHI) (Rosenthal and Strange, 2003). We also use the Herfindahl measure to capture the competitive degree of a specific sector located within a specific administrative unit. Consistent with the competition literature (Nickell, 1996; Earle and Estrin, 2003; Syverson, 2004; Schmitz Jr, 2005) , we confirm the productivity benefits stemming from competition.

Since the works of Aschauer (1989) and Munnell (1992), several authors attempt to establish a relationship between public infrastructure spending and economic growth or productivity using aggregate data (Holtz-Eakin, 1992; Hulten and Schwab, 1991; Pereira and Andraz, 2003; Fernald, 1999). It is not surprising to see that public investments have been widely used by decision makers to foster economic growth, especially during the economic downturn (see Ligthart and Suarez, 2011 for a survey on public capital). The extensive research findings on such linkage since late 1980s, however, are characterised by a wide range of estimates. Taking public spending on transportation for example, Melo, Graham, and Brage-Ardao (2013) in their survey paper conducting a meta-analysis of the empirical evidence on the output elasticity of infrastructure, based on 563 estimates obtained from 33 studies, indicates that the variances are attributed to the model specification, the aggregation of data, the data type (time series or panel), industrial scope, the level of study (national or sub-national), and whether appropriate instruments are adopted to avoid endogeneity. A few Chinese sub-national level studies also contribute to this large strand of literature. For example, Vijverberg, Fu, and Vijverberg (2011), using province-level data from 1993 to 2003, estimate cost function models of production in industrial enterprises, finding that on average public infrastructure contributes 2-3% points to the growth in labor productivity among these enterprises. Demurger (2001), using panel data from a sample of 24 Chinese provinces (excluding municipalities) throughout 1985-1998, argues that infrastructure endowment did account significantly for observed differences in growth performance across provinces. With a spatial dimension, the spillover effect of transport infrastructure is identified for China by Yu, De Jong, Storm, and Mi (2013). Using panel data for the Spanish provinces over the period 1985–2004, Gomez-Antonio and Fingleton (2012) find consistent evidence that productivity depends directly on the public capital stock endowment of each province, but negative spillover effects are predicted from changes in capital stock in neighboring provinces. Different from the current stock of literature, we are among the first to use micro-level data to examine the spillover effect of public assets in China across rural counties and urban districts. It is interesting to note that benefits of public works tend to be localized and the productivity boosting effect of investment from the public sector is confirmed on firms in the electronics sector. Compared with existing studies, the marginal effects of public investment in neighbouring districts and counties are much smaller in size, although being highly significant. Furthermore, we show that larger administrative units are systematically able to extract more benefits from the political process than are smaller ones.

By controlling for the five TFP shifters–R&D, export, specialization, competition and local public spending, and constructing two spatial lags, our baseline IV-2SLS estimators in the fixed effect model indicate that the spatial interactions among firms in the electronics sector arising from "agglomeration" are reciprocal. Four-percent (one-percent) of the productivity growth of a specific firm stems from a ten-percent productivity increase of its close (faraway) neighbors located in the same (neighboring) jurisdiction. In contrast with Baltagi, Egger, and Kesina (2015), our findings suggest that the productivity transmission process in space attenuates with distance. These results are shown to be robust to the change in the spatial lag using the distance measure as well as adoption of a more efficient GMM estimation strategy. By dividing the administrative units into densely populated urban district with high quality public infrastructure and the less concentrated county or county-level cities, the base line model behaves in different ways. The productivity spillover effect from the type-II neighbors is significant and larger in size for the urban areas. Intuitively, the urban districts as the traditionally economic and political centers have closer ties with their neighbors. In addition, through conducting similar tests sorting 2866 administrative units by area, it is shown that the distance matters. The between-region spillover effect in productivity growth turns out to be insignificant for larger areas. The only exception is found when we interact the spatial lag with both administrative type (urban or county) and size of the location (small or large). It seems that counties benefit more from their type-II neighbors than the urban. By and large, we find out that both regional type and size have influences on the productivity transmission in space.

The next Section elaborates on the empirical methodology in which we compare a fixed effect model specification with the random effect and discuss the analytical difficulty brought by the noncommutative property of the within-transformation and the spatial lag matrix to use the same specification with both effects. Section 6 describes the three data sets we have used–CASIF, Chinese Prefecture City and County Fiscal Statistical Material, and the encoded Chinese administrative unit map together with some auxiliary GB documents. The empirical results are presented in Section 7 and the last section provides concluding remarks.

5 Empirical methodology

In this Section, we first estimate firm-level productivity and then explore the sources of firm productivity growth. Since the panel data is unbalanced, care must be undertaken because spatial regression models based on balanced data do not extend flawlessly to unbalanced panels.

5.1 Estimation of total factor productivity

In this study, we adopt Levinsohn and Petrin's (2003) two-step method to estimate TFP. The Levinsohn-Petrin (hereafter LP) method enables us to obtain unbiased estimates of total factor productivity (TFP), even if the variable inputs are endogenous to market conditions or other time-varying unobservables that affect productivity.[11] We assume the following Cobb-Douglas production function

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \eta_{it}, \tag{1}$$

where y_{it}, l_{it}, k_{it} are respectively the logarithm of value added, labor input, capital stock of firm i in year t. ω_{it} is the total factor productivity, and η_{it} is the error term. The LP method starts with estimating the 1$^{\text{st}}$-stage least-squares regression

$$y_{it} = \beta_l l_{it} + \phi\left(k_{it}, m_{it}\right) + \eta_{it}, \tag{2}$$

where $\phi(\cdot)$ is a three-order polynomial of capital k_{it} and intermediate input m_{it}. An estimate of $\phi\left(k_{it}, m_{it}\right)$ is then constructed as

$$\hat{\phi}_{it} = y_{it} - \hat{\beta}_l l_{it}. \tag{3}$$

The 2$^{\text{nd}}$-stage of the LP procedure estimates β_k through the following nonlinear regression

$$y_{it} - \hat{\beta}_l l_{it} - \beta_k k_{it} = \psi\left(\hat{\phi}_{it-1} - \beta_k k_{it-1}\right) + \eta_{it}, \tag{4}$$

where $\psi(\cdot)$ is another polynomial of order three. In practice, (4) is estimated by minimizing the sum of squared errors. Once $\hat{\beta}_l$ and $\hat{\beta}_k$ are estimated, the predicted value of ω_{it} is given by

$$\hat{\omega}_{it} = y_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it}, \tag{5}$$

which is our measure of TFP.

[11] Our choice of the LP method over that of Olley and Pakes (1996) is largely based on data concerns. The OP estimator proxies productivity by firm's investment decision and state of exit. However, investment is known to have limitations and may not be applicable in general (Levinsohn and Petrin, 2003). More seriously, our data does not provide information on investment. Thus it must be derived from the capital stock. Data on capital stock prepared by the NBS are infamous for being systematically biased. Although various methods have been developed to estimate the true capital stock (Brandt, Van Biesebroeck, and Zhang, 2012), they all rely on some subjective parameter, thus may introduce substantial noise to the final result. Information on firm's exit is also problematic since it does not truly reflect a change in the operational status, thus is not a valid proxy. With the LP method, we are able to circumvent all these obstacles.

5.2 The spatial fixed effects model

We assume the following general nested spatial model with two autoregressive terms:

$$\begin{aligned} \mathbf{y}_t &= \lambda \mathbf{W}_t \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \alpha \boldsymbol{\iota}_t + \boldsymbol{\mu}_t + \mathbf{u}_t, \\ \mathbf{u}_t &= \rho \mathbf{M}_t \mathbf{u}_t + \boldsymbol{\epsilon}_t, \end{aligned} \tag{6}$$

where $t = 1, \cdots, T$ denote the time periods. Let the number of observations in year t be N_t, the dependent variable $\mathbf{y}_t$ in (6) is an $N_t \times 1$ vector of firm-level TFP estimates in year t. $\mathbf{W}_t$ and $\mathbf{M}_t$ are two $N_t \times N_t$ maximum row-normalized spatial weight matrices with zeros in the main diagonal. Because the number of observations varies with time, both $\mathbf{W}_t$ and $\mathbf{M}_t$ are time-varying. $\mathbf{X}_t$ is an $N_t \times K$ matrix of K exogenous regressors. $\boldsymbol{\iota}_t$ is an $N_t \times 1$ vector of all ones. $\boldsymbol{\mu}_t$ is an $N_t \times 1$ vector of individual fixed effects. If the same firm exists in period t and t', the column entries of $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_{t'}$ that correspond to the same firm must be identical in value. Finally, the error term $\mathbf{u}_t$ is assumed to be generated by a spatial autoregressive process with i.i.d. disturbances $\boldsymbol{\epsilon}_t$ whose mean is zero and variance is σ_ϵ^2.[12]

Stacking the equations over time periods, we can transform (6) into its panel representation

$$\begin{aligned} \mathbf{y} &= \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \alpha \boldsymbol{\iota} + \boldsymbol{\mu} + \mathbf{u}, \\ \mathbf{u} &= \rho \mathbf{M} \mathbf{u} + \boldsymbol{\epsilon}. \end{aligned} \tag{7}$$

Here $N = \sum_{t=1}^{T} N_t$ is the total number of observations. $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \cdots, \mathbf{y}_T')'$, and other vectors (including $\mathbf{X}$) are defined similarly. $\mathbf{W} = \operatorname{diag}(\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_T)$ is an $N \times N$ block-diagonal matrix with $\mathbf{W}_t$, $t = 1, \cdots, T$ on the diagonal. $\mathbf{M}$ is constructed in the same way.

Clearly, (7) implies

$$E(\mathbf{u}\mathbf{u}') = \sigma_\epsilon^2 (\mathbf{I}_N - \rho \mathbf{M})^{-1} (\mathbf{I}_N + \rho \mathbf{M}')^{-1} = \sigma_\epsilon^2 \left(\mathbf{I}_N - \rho(\mathbf{M} + \mathbf{M}') - \rho^2 \mathbf{M}'\mathbf{M}\right)^{-1}, \tag{8}$$

and

$$E((\mathbf{W}\mathbf{y})\mathbf{u}') = E\left(\mathbf{W}(\mathbf{I}_N - \lambda \mathbf{W})^{-1} \mathbf{u}\mathbf{u}'\right) = \mathbf{W}(\mathbf{I}_N - \lambda \mathbf{W})^{-1} E(\mathbf{u}\mathbf{u}') \neq \mathbf{0}.$$

Therefore, we have both endogeneity and non-spherical disturbances. To obtain consistent estimates of the structural parameters, we can find instruments for the RHS endogenous variable $\mathbf{W}\mathbf{y}$. It remains to circumvent the incidental parameter problem by taking the within transformation of (7) , if the fixed effects themselves are not the interest of the study. For convenience, let's denote the matrices of within and between transformations by $\mathbf{Q}_0$ and $\mathbf{Q}_1$, respectively.[13] Since $\mathbf{Q}_0 \boldsymbol{\mu} = \mathbf{0}$, the within transformation eliminates the fixed effects from (7), so that we have

$$\mathbf{Q}_0 \mathbf{y} = \lambda \mathbf{Q}_0 \mathbf{W} \mathbf{y} + \mathbf{Q}_0 \mathbf{X} \boldsymbol{\beta} + \mathbf{Q}_0 \mathbf{u}. \tag{10}$$

[12] $\boldsymbol{\epsilon}_t$ from different cross sections are also assumed to be independent.

[13] $\mathbf{Q}_0$ and $\mathbf{Q}_1$ do not have neat matrix representations given the way we order observations in (7).

Despite the autoregressive structure in $\mathbf{u}$, the error term in (10) has zero mean, and the expectation of the RHS endogenous variable $\mathbf{Q}_0\mathbf{W}\mathbf{y}$ turns out to be

$$\begin{aligned} E\left(\mathbf{Q}_0\mathbf{W}\mathbf{y}\right) &= \mathbf{Q}_0\mathbf{W}\left(\mathbf{I}_N - \lambda\mathbf{W}\right)^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} + \alpha\boldsymbol{\iota}\right) \\ &= \mathbf{Q}_0\sum_{k=0}\lambda^k\mathbf{W}^{k+1}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} + \alpha\boldsymbol{\iota}\right), \end{aligned} \tag{11}$$

which suggests that $\mathbf{Q}_0\mathbf{W}\mathbf{y}$ in (10) can be instrumented by[14]

$$\mathbf{G}_0 = \big(\mathbf{Q}_0\mathbf{X}, \mathbf{Q}_0\mathbf{W}\mathbf{X}, \mathbf{Q}_0\mathbf{W}^2\mathbf{X}, \cdots\big). \tag{12}$$

In this way, the structural parameters $\boldsymbol{\delta} = (\lambda, \boldsymbol{\beta}', \alpha)'$ can be consistently estimated by 2SLS, which estimator we denote by $\hat{\boldsymbol{\delta}}_W$.

The above is the procedure proposed by Kelejian and Prucha (1998). $\hat{\boldsymbol{\delta}}_W$ is consistent as far as $\mathbf{u}$ is orthogonal to the exogenous regressors $\mathbf{X}$. It remains to find a proper standard error for this estimator. Let's denote $(\mathbf{Q}_0\mathbf{W}\mathbf{y}, \mathbf{Q}_0\mathbf{X})$ by $\mathbf{Z}_0$ and the projection matrix onto $\mathbf{G}_0$ by $\mathbf{P}_{\mathbf{G}_0}$, then

$$E\left(\hat{\boldsymbol{\delta}}_W - \boldsymbol{\delta}\middle|\mathbf{X}\right) = \left(\frac{\mathbf{Z}_0'\mathbf{P}_{\mathbf{G}_0}\mathbf{Z}_0}{N}\right)^{-1}\frac{\mathbf{Z}_0'\mathbf{P}_{\mathbf{G}_0}\mathbf{u}}{N}. \tag{13}$$

It is clear from (8) that there are heteroskedasticity and spatial correlations in $\mathbf{u}$. The conventional HAC standard errors are incapable of modeling such correlations. In this study we obtain the standard errors by cluster bootstrapping, where each distinct firm is treated as a cluster. Note that $\hat{\boldsymbol{\delta}}_W$ is simply a one-step GMM estimator using equal weights for the moment conditions. The two-step or the iterated GMM estimator is theoretically more efficient. Again, bootstrapped standard errors are preferred in this case.

5.3 The spatial random effects model

A competing model for (7) is the random-effects specification

$$\begin{aligned} \mathbf{y} &= \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \alpha\boldsymbol{\iota} + \mathbf{u}, \\ \mathbf{u} &= \rho\mathbf{M}\mathbf{u} + \boldsymbol{\mu} + \boldsymbol{\epsilon}, \end{aligned} \tag{7$'$}$$

If we permute (7) so that observations are ordered first by individual i then by time t, then

$$\begin{aligned} \mathbf{Q}_1' &= \operatorname{diag}\left(\frac{1}{T_1}\boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}', \frac{1}{T_2}\boldsymbol{\iota}_{T_2}\boldsymbol{\iota}_{T_2}', \cdots, \frac{1}{T_n}\boldsymbol{\iota}_{T_n}\boldsymbol{\iota}_{T_n}'\right), \\ \mathbf{Q}_0' &= \mathbf{I}_N - \mathbf{Q}_1, \end{aligned} \tag{9}$$

in which T_i denotes the number of time periods in which individual i is observed. Therefore, the matrix representation of $\mathbf{Q}_0$ and $\mathbf{Q}_1$ can be obtained with a proper permutation of (9).

[14] Although a similar construction based on $\boldsymbol{\mu}$ can also be used as instruments, it is infeasible since $\boldsymbol{\mu}$ is unknown at this stage.

where the individual effect $\boldsymbol{\mu}$ is assumed to be uncorrelated with $\mathbf{X}$. The FG2SLS procedure of Kelejian and Prucha (1998) or Mutl and Pfaffermayr (2011) can be easily adapted to unbalanced panels under the random effects assumption. Despite the fact that $\mathbf{Q}_0$ does not commute with $\mathbf{M}$, and that $\boldsymbol{\mu}$ in $\mathbf{u}$ does not vanish after the within transformation, $\mathbf{Q}_0\mathbf{u}$ as a whole is uncorrelated with $\mathbf{Q}_0\mathbf{X}$. Thus, the 2SLS estimator $\hat{\boldsymbol{\delta}}_W$ obtained from (10) remains consistent. Hereby the residuals

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\lambda}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\alpha}\boldsymbol{\iota} \tag{14}$$

are consistent estimates of the error terms in (7′).

Although $\hat{\boldsymbol{\delta}}_W$ is consistent, they are not efficient since the error terms $\mathbf{Q}_0\mathbf{u}$ in (10) are non-spherical. A GLS transformation of (10) or (7) followed by another least-squares will restore efficiency. Thus, the next step is to estimate the parameter ρ in the error term $\mathbf{u}$. Since $\boldsymbol{\epsilon}$ is assumed to be spherical and all diagonal elements of $\mathbf{M}$ are zero, we have

$$\begin{aligned}
&E\left(\boldsymbol{\epsilon}'\mathbf{Q}_0\boldsymbol{\epsilon}\right) = \operatorname{tr}\left(\mathbf{Q}_0\right)\sigma_\epsilon^2 = (N-n)\sigma_\epsilon^2,\\
&E\left(\boldsymbol{\epsilon}'\mathbf{Q}_0\mathbf{M}\mathbf{Q}_0\boldsymbol{\epsilon}\right) = \operatorname{tr}\left(\mathbf{Q}_0\mathbf{M}\right)\sigma_\epsilon^2 = 0,\\
&E\left(\boldsymbol{\epsilon}'\mathbf{Q}_0\mathbf{M}'\mathbf{M}\mathbf{Q}_0\boldsymbol{\epsilon}\right) = \operatorname{tr}\left(\mathbf{Q}_0\mathbf{M}'\mathbf{M}\right)\sigma_\epsilon^2 = \operatorname{tr}\left(\operatorname{diag}\left(\mathbf{Q}_0\right)\operatorname{diag}\left(\mathbf{M}'\mathbf{M}\right)\right)\sigma_\epsilon^2,\\
&E\left(\mathbf{u}'\mathbf{Q}_1\mathbf{u}\right) = N\sigma_\mu^2 + n\sigma_\epsilon^2.
\end{aligned} \tag{15}$$

Here n denotes the number of distinct individuals (firms) in the unbalanced panel, $\operatorname{diag}\left(\mathbf{Q}_0\right)$ denotes the diagonal matrix of $\mathbf{Q}_0$, and $\operatorname{diag}\left(\mathbf{M}'\mathbf{M}\right)$ is defined similarly.

Since $\mathbf{Q}_0\boldsymbol{\epsilon} = \mathbf{Q}_0\left(\mathbf{I}_N - \rho\mathbf{M}\right)\mathbf{u}$, (15) lead to the following moment conditions:

$$\begin{aligned}
&\hat{\mathbf{u}}'\left(\mathbf{I}_N - \rho\mathbf{M}'\right)\mathbf{Q}_0\left(\mathbf{I}_N - \rho\mathbf{M}\right)\hat{\mathbf{u}} = (N-n)\sigma_\epsilon^2,\\
&\hat{\mathbf{u}}'\left(\mathbf{I}_N - \rho\mathbf{M}'\right)\mathbf{Q}_0\mathbf{M}\mathbf{Q}_0\left(\mathbf{I}_N - \rho\mathbf{M}\right)\hat{\mathbf{u}} = 0,\\
&\hat{\mathbf{u}}'\left(\mathbf{I}_N - \rho\mathbf{M}'\right)\mathbf{Q}_0\mathbf{M}'\mathbf{M}\mathbf{Q}_0\left(\mathbf{I}_N - \rho\mathbf{M}\right)\hat{\mathbf{u}} = \operatorname{tr}\left(\operatorname{diag}\left(\mathbf{Q}_0\right)\operatorname{diag}\left(\mathbf{M}'\mathbf{M}\right)\right)\sigma_\epsilon^2,\\
&\hat{\mathbf{u}}'\left(\mathbf{I}_N - \rho\mathbf{M}'\right)\mathbf{Q}_1\left(\mathbf{I}_N - \rho\mathbf{M}\right)\hat{\mathbf{u}} = N\sigma_\mu^2 + n\sigma_\epsilon^2.
\end{aligned} \tag{16}$$

The three unknowns ρ, σ_ϵ^2, and σ_μ^2 can be estimated from (16) by GMM. Let's denote the estimates by $\left(\hat{\rho}, \hat{\sigma}_\epsilon^2, \hat{\sigma}_\mu^2\right)$, with which we can perform the FGLS transformation on (7):

$$\begin{aligned}
\boldsymbol{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}_N - \hat{\rho}\mathbf{M}\right)\mathbf{y} = \lambda\boldsymbol{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}_N - \hat{\rho}\mathbf{M}\right)\mathbf{W}\mathbf{y} &+ \boldsymbol{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}_N - \hat{\rho}\mathbf{M}\right)\mathbf{X}\boldsymbol{\beta}\\
&+ \alpha\boldsymbol{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}_N - \hat{\rho}\mathbf{M}\right)\boldsymbol{\iota} + \boldsymbol{\nu}.
\end{aligned} \tag{17}$$

$\boldsymbol{\Omega}^{-\frac{1}{2}}$ in (17) is the conventional Cochrane-Orcutt transformation for unbalanced panels. For any variable ξ,

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\xi_{it} = \xi_{it} - \frac{\hat{\sigma}_\epsilon}{\left(T_i\hat{\sigma}_\mu^2 + \hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}}\bar{\xi}_{i.}, \tag{18}$$

where T_i is the number of observations pertinent to individual i (Baltagi, Egger, and

Kesina, 2015).

The last step of the procedure is a 2SLS regression on (17) with the RHS endogenous regressor $\mathbf{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}_N-\hat{\rho}\mathbf{M}\right)\mathbf{W}\mathbf{y}$ instrumented by

$$\begin{aligned}\mathbf{G}_1=\big(&\mathbf{Q}_0\mathbf{X},\mathbf{Q}_0\mathbf{W}\mathbf{X},\mathbf{Q}_0\mathbf{W}^2\mathbf{X}\cdots,\mathbf{Q}_0\mathbf{M}\mathbf{X},\mathbf{Q}_0\mathbf{M}\mathbf{W}\mathbf{X},\mathbf{Q}_0\mathbf{M}\mathbf{W}^2\mathbf{X}\cdots,\\&\mathbf{Q}_1\mathbf{X},\mathbf{Q}_1\mathbf{W}\mathbf{X},\mathbf{Q}_1\mathbf{W}^2\mathbf{X}\cdots,\mathbf{Q}_1\mathbf{M}\mathbf{X},\mathbf{Q}_1\mathbf{M}\mathbf{W}\mathbf{X},\mathbf{Q}_1\mathbf{M}\mathbf{W}^2\mathbf{X}\cdots,\\&\mathbf{Q}_0\mathbf{W}\boldsymbol{\iota},\mathbf{Q}_0\mathbf{W}^2\boldsymbol{\iota}\cdots,\mathbf{Q}_0\mathbf{M}\boldsymbol{\iota},\mathbf{Q}_0\mathbf{M}\mathbf{W}\boldsymbol{\iota},\mathbf{Q}_0\mathbf{M}\mathbf{W}^2\boldsymbol{\iota}\cdots,\\&\boldsymbol{\iota},\mathbf{Q}_1\mathbf{W}\boldsymbol{\iota},\mathbf{Q}_1\mathbf{W}^2\boldsymbol{\iota}\cdots,\mathbf{Q}_1\mathbf{M}\boldsymbol{\iota},\mathbf{Q}_1\mathbf{M}\mathbf{W}\boldsymbol{\iota},\mathbf{Q}_1\mathbf{M}\mathbf{W}^2\boldsymbol{\iota}\cdots,\big),\end{aligned}\tag{19}$$

which is the optimal set of instruments in the random effects setup.

Under certain assumptions, Mutl and Pfaffermayr (2011) show that the FG2SLS estimator of the random effects model has the expected asymptotic distribution. If we denote $\left(\mathbf{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}-\rho\mathbf{M}\right)\mathbf{W}\mathbf{y},\mathbf{\Omega}^{-\frac{1}{2}}\left(\mathbf{I}-\rho\mathbf{M}\right)\mathbf{X}\right)$ by $\widetilde{\mathbf{Z}}$, and the projection matrix onto $\mathbf{G}_1$ by $\mathbf{P}_{\mathbf{G}_1}$, then the FG2SLS random effects estimator $\hat{\boldsymbol{\delta}}_R$ has the asymptotic distribution

$$\hat{\boldsymbol{\delta}}_R\xrightarrow{a}N\left(\boldsymbol{\delta},\frac{\sigma_\epsilon^2}{N}\left(\frac{\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}}{N}\right)^{-1}\left(\frac{\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}}{N}\right)\left(\frac{\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}}{N}\right)^{-1}\right).\tag{20}$$

This suggests that the variance-covariance matrix of $\widetilde{\boldsymbol{\delta}}_R$ can be estimated by

$$\hat{\sigma}_\epsilon^2\left(\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}\right)^{-1}\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}\left(\widetilde{\mathbf{Z}}'\mathbf{P}_{\mathbf{G}_1}\widetilde{\mathbf{Z}}\right)^{-1},\tag{21}$$

where $\hat{\sigma}_\epsilon^2$ is obtained from the second stage of the procedure.

5.4 Issues with unbalanced panels

Our specification of the fixed effects model (7) is inconsistent with that of the random effects model (7′). In the FG2SLS literature (e.g. Kelejian, Prucha, and Yuzefovich, 2004; Mutl and Pfaffermayr, 2011), both models are given by (7′), i.e., the individual effects are assumed to be a component of the disturbances. This specification allows a similar feasible GLS transformation on the fixed effects model. The 2SLS estimator obtained from the transformed equation is more efficient than $\hat{\boldsymbol{\delta}}_W$. More importantly, the unified treatment facilitates a subsequent Hausman test. Such advantages, however, rely heavily on the commutativity of the within and the spatial lag transformations, which is automatically satisfied if the panel is balanced and if the spatial weight matrices are time-invariant. Without commutativity, this advantage becomes an analytical burden. To see this, let's note that the within-transformation of (7′) gives

$$\mathbf{Q}_0\mathbf{y}=\lambda\mathbf{Q}_0\mathbf{W}\mathbf{y}+\mathbf{Q}_0\mathbf{X}\boldsymbol{\beta}+\mathbf{Q}_0\left(\mathbf{I}-\rho\mathbf{M}\right)^{-1}\left(\boldsymbol{\mu}+\boldsymbol{\epsilon}\right).\tag{10′}$$

Since $\mathbf{Q}_0\left(\mathbf{I}_N-\rho\mathbf{M}\right)^{-1}\neq\left(\mathbf{I}_N-\rho\mathbf{M}\right)^{-1}\mathbf{Q}_0$, the individual effects $\boldsymbol{\mu}$ do not vanish. Under the fixed effects assumption, the error component $\mathbf{Q}_0\left(\mathbf{I}-\rho\mathbf{M}\right)^{-1}\left(\boldsymbol{\mu}+\boldsymbol{\epsilon}\right)$ is again correlated with $\mathbf{X}$, so the within estimator becomes inconsistent.

By moving the individual effects from the error component to the structural model, (7) ensures the consistency of $\hat{\boldsymbol{\delta}}_W$. With (7), however, the three-stage procedure is no longer feasible. This is because the estimation of ρ is based on the residuals from the first-stage 2SLS, namely

$$\hat{\mathbf{u}}_0 = \widehat{\mathbf{Q}_0\mathbf{u}} = \mathbf{Q}_0\left(\mathbf{y} - \hat{\lambda}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\alpha}\boldsymbol{\iota}\right). \tag{22}$$

If the spatial weights $\mathbf{M}_t$ are time-invariant, then the within transformation $\mathbf{Q}_0$ and the spatial lag operation $\mathbf{M}$ are commutative, so that

$$\left(\mathbf{I} - \rho\mathbf{M}\right)\mathbf{Q}_0\mathbf{u} = \mathbf{Q}_0\left(\mathbf{I} - \rho\mathbf{M}\right)\mathbf{u} = \mathbf{Q}_0\boldsymbol{\epsilon}. \tag{23}$$

Therefore, $\left(\mathbf{I} - \rho\mathbf{M}\right)\widehat{\mathbf{Q}_0\mathbf{u}}$ can be used to construct the moment conditions regarding $\mathbf{Q}_0\boldsymbol{\epsilon}$. Without commutativity, however, (23) is invalid and $\left(\mathbf{I} - \rho\mathbf{M}\right)\widehat{\mathbf{Q}_0\mathbf{u}}$ becomes an estimate of $\left(\mathbf{I} - \rho\mathbf{M}\right)\mathbf{Q}_0\left(\mathbf{I} - \rho\mathbf{M}\right)^{-1}\boldsymbol{\epsilon}$, which contains the unknown parameter ρ.

Even if we are given a consistent estimator of ρ, the time-varying spatial weights remain an obstacle to GLS estimation. To see this point, let's note that in the transformed equation

$$\left(I - \rho\mathbf{M}\right)\mathbf{y} = \lambda\left(I - \rho\mathbf{M}\right)\mathbf{W}\mathbf{y} + \left(I - \rho\mathbf{M}\right)\mathbf{X}\beta + \left(I - \rho\mathbf{M}\right)\left(\alpha\boldsymbol{\iota} + \boldsymbol{\mu}\right) + \boldsymbol{\epsilon},$$

$\left(I - \rho\mathbf{M}\right)\boldsymbol{\mu}$ must be eliminated before estimating the structural parameters. In balanced panel models, this is done by a within transformation because $\mathbf{M}$ commutes with $\mathbf{Q}_0$. Without commutativity, the within transformation is bound to fail.

The above discussion reveals the critical constraint imposed by unbalanced panels on the fixed effects model. In order to obtain consistent estimates of the structural parameters from least squares, one must choose (7) over (7′). By doing so, one has to forfeit the efficiency gains from the FG2SLS procedure. Nevertheless, the work by Kelejian, Prucha, and Yuzefovich (2004) shows that such efficiency gains, if any, could be small in magnitude with even a moderate sample size.

Unbalanced panels also introduce minor changes to the FG2SLS procedure for the random effects model. Because $\mathbf{Q}_0$ does not commute with $\mathbf{M}$, the moment conditions in (15) differ from their balanced-panel counterparts. For a similar reason, (19) now consists of more instruments.

Although we use different specifications for the fixed and random effects models, the fixed effects estimator $\hat{\boldsymbol{\delta}}_W$ coincides with the first-stage within estimator of the random effects model. Given specification (7′), both $\hat{\boldsymbol{\delta}}_W$ and $\hat{\boldsymbol{\delta}}_R$ are consistent under the random effects assumption that $E\left(\boldsymbol{\mu}|\mathbf{X}\right) = \mathbf{0}$, but the latter is more efficient. Therefore, we can design a Hausman test by comparing $\hat{\boldsymbol{\delta}}_W$ and $\hat{\boldsymbol{\delta}}_R$. If the random effects assumption is rejected, we shall estimate the fixed effects model (7) and base our inference on this alternative specification.

6 Data description and measurement

We obtain firm-level data on accounting and financial variables from the CASIF. Data on accounting and financial variables, including total value added, net value of fixed assets, total employment, total intermediate inputs, etc., enable us to estimate firm-level productivity and construct proxies for firm characteristics. The data set also provides location information of each firm, with which a serious spatial analysis is possible.

The geo-data are compiled from multiple sources. Our major reference is the official *Code Book of Administrative Divisions* prepared by the National Bureau of Statistics. The code book, published as the National Standard GB/T 2260, assigns six-digit administrative codes to over 3,000 administrative divisions in four levels of the hierarchy: province, prefecture, sub-prefecture, district/county. As of 2007, there were 2866 administrative units in the district/county level. With the code book, we are able to identify the location (district/county) of each establishment. We acquire the geographic data of the administrative units from a commercial source. These include the coordinates of the administrative centers and a shape file of administrative boundaries. The geo-data enables us to construct spatial neighborhood relations among administrative units either by spatial distance or by contiguity.

Our study also employs data on the budgetary expenditure of local governments in the district/county level. The data is extracted from *Chinese Prefecture City and County Statistical Material (Various issues 1999-2007)* compiled by the Ministry of Finance. It is then merged into the main data set by matching administrative codes. There are two issues with the fiscal data. First, it uses a different administrative division than the one specified by the official code book. Data loss is inevitable when we merge data. Nevertheless, we are able to retain most of the administrative units at the district/county level.[15] Second, the expenditure categories were revised twice during the study period, first in 2003 and then in 2007. Consequently, the expenditure categories in different issues of the yearbook are not compatible, except for total budgetary expenditure.

In this study, we focus on the sector of electric apparatus (industry code 3900) over the period of 1999-2007. The choice of this sector is based on the following concerns: First, the sector is technologically intensive. Agglomeration effects on productivity may be more pronounced thus easier to detect for this sector. Second, this sector provides a large sample size (over 99,000), with over 27,596 firms located in 1629 districts/counties. The presence of the sector in a large number of contiguous administrative units allows us to conduct spatial analysis without a heavy penalty of data loss.[16] The spatial distribution of firms and employment in the sector are shown by Figures (1-2).[17] A pattern of agglomeration is evident in these graphs. Firms and

[15]The number of dropouts ranges from 124 in 1999 to 34 in 2006. The city of Shenzhen was treated as one piece by the yearbook until 2007. The city is known to be a major manufacturing hub in the Pearl River Delta, hosting a large number of firms. In order to retain these observations in our sample, we treat the five urban districts of Shenzhen as one administrative unit.

[16]We removed observations that are "islands."

[17]The figures use three year averages (log-transformed values) over 2005-2007.

employment cluster on the eastern coast and inland industrial centers, while the vast areas in the west are unoccupied. This observation suggests a strong spatial linkage in the location choice of firms, and likely spatial interactions between firms when they are close.

The estimation of (1) requires firm-level data on value added, inputs in capital, labor, and intermediate goods. In this study, y_{it} is measured by total value added of the firm, l_{it} by annual average number of employed personnel, k_{it} by annual average value of net fixed assets, and m_{it} by value of intermediate inputs. Data on these variables are available from the CASIF with a few exceptions. Total value added is missing in 2001, 2002, and 2004, thus must be derived from other variables by accounting identities. The 2001 and 2002 values are computed as

$$\begin{aligned}\text{total value added} = {} & \text{gross industrial output - value of intermediate inputs} \\ & + \max\{0, \text{value added tax payable}\},\end{aligned}$$

while those of 2004 are recovered from

$$\begin{aligned}\text{total value added} = {} & \text{revenue from principal business} + \text{increase in inventory} \\ & \text{- value of intermediate inputs} + \max\{0, \text{value added tax payable}\}.\end{aligned}$$

We plot the weighted (by employment) average TFP of each district/county in Figure (3). Compared to Figures (1-2), the spatial pattern of TFP is less clear, partly because individual heterogeneity is smoothed out by taking the average.[18] Nevertheless, in small clusters, such as the metro areas of Chengdu, Guangzhou, and Wuhan, we do observe the spatial gradient of TFP declining from the center to the periphery.

The basic geographic unit in our data is an urban district or a county. The location of a firm is identified by the district/county in which it operates, but we have no further locational information within the district/county. We assume that firms in the same district/county are all located at the administrative center. In this regard, a firm has two types of neighbors: those in the same district/county and those in neighboring districts/counties. Thus, we introduce two spatial weight matrices to the SAR model (6):

$$\mathbf{W}_{1t} y_{ikt} = \sum_{\substack{j \in I_t(k) \\ j \neq i}} l_{jkt} y_{jkt} \Big/ \sum_{\substack{j \in I_t(k) \\ j \neq i}} l_{jkt}, \tag{24}$$

$$\mathbf{W}_{2t} y_{ikt} = \sum_{\substack{j \in I_t(k') \\ k' \in N(k)}} l_{jk't} y_{jk't} \Big/ \sum_{\substack{j \in I_t(k') \\ k' \in N(k)}} l_{jk't}. \tag{25}$$

Here the subscripts i and j denote firms, k and k' denote districts/counties, and t

[18] A few studies (e.g. Hu, Xu, and Yashiro, 2015) show that firm type, including ownership structure and size, is correlated with productivity. If a firm type is inproportionally high or low in a district/county, the average TFP will be biased compared to those of neighbors.

denotes time. $I_t(k)$ is the set of all firms located in district/county k in year t, and $N(k)$ is the set of all neighboring districts/counties of district/county k. y is any variable to be weighted, and l_{ikt} is the employment of firm i in district/county k and year t. The construction is based on the premise that larger firms (measured by employment) exert stronger influence on their neighbors than smaller ones. Contiguous (Rook style) districts/counties are treated as neighbors. We also use an alternative definition based on distance. Two districts/counties are regarded as neighbors if their administrative centers are within 50 kilometers in great circle distance.[19]

Both $\mathbf{W}_{1t}$ and $\mathbf{W}_{2t}$ are maximum row normalized, and they have zeros in the main diagonal. It is easy to see $\mathbf{W}_{1t}\mathbf{W}_{2t} = \mathbf{W}_{2t}$. This property helps to alleviate the computation burden in the regression stage. For simplicity, we use a single AR term in the error component, i.e., we assume $\mathbf{u} = \rho\mathbf{W}_1\mathbf{u} + \boldsymbol{\epsilon}$ in (7) and $\mathbf{u} = \rho\mathbf{W}_1\mathbf{u} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$ in (7′).

The spatial model (6) considers two types of exogenous variables: firm idiosyncrasies that have no effect on other firms, and market conditions that impact not only local firms, but also likely firms in neighboring districts/counties. The literature (Sheng and Song, 2013; Hu, Xu, and Yashiro, 2015; Baltagi, Egger, and Kesina, 2015) has identified multiple firm-level characteristics that are correlated with productivity, including ownership structure, size of the firm, years of operation, R&D activity, and participation in the international market. Our current study specifies individual effects. Thus, only time-varying factors can be properly estimated by the model. We use two variables to proxy R&D (rd) and export (ex) activities. They are respectively measured by the share of new merchandise in gross output and the fraction of gross sales that are exported. We consider both the local market condition and the institutional environment: specialization (*spec*), competition (*comp*), and public spending (*pub*). According to Marshall's (1890) hypothesis, a city benefits from specialization because of spillovers between firms in the same industry. Our measure follows that of Glaeser, Kallal, Scheinkman, and Shleifer (1992), i.e.,

$$spec_{kt} = \frac{\dfrac{\text{sectoral employment in area } k \text{ and year } t}{\text{total industrial employment in area } k \text{ and year } t}}{\dfrac{\text{sectoral employment in China and year } t}{\text{total industrial employment in China and year } t}}$$

Porter (1990) argues that competition among local firms boosts productivity. Instead of measuring the average firm size (Glaeser, Kallal, Scheinkman, and Shleifer, 1992; Rosenthal and Strange, 2003), we use the Herfindahl-Hirschman index of sectoral employment in the district/county. Using the notations from (24), we define

$$comp_{kt} = \sum_{i \in I_t(k)} \left(\frac{l_{ikt}}{\sum_{i \in I_t(k)} l_{ikt}} \right)^2.$$

[19]The choice of the 50-kilometer cutoff value is based on the observation that the mean distance between contiguous neighbors is 64 kilometers, while the median is 47 kilometers. In this way, the estimate of λ_2 will have a similar interpretation as in the contiguity case.

The HHI take value between zero and unity. A smaller value indicates stronger competition. Because the HHI decreases in the number of firms even if the distribution of employment among firms remains unchanged, it also proxies the number of firms, which has also been proposed as a measure of agglomeration (Hu, Xu, and Yashiro, 2015). Although the relationship between public investment and productivity has been studied for long (Aschauer, 1989; Fernald, 1999; Vijverberg, Fu, and Vijverberg, 2011), it remains a missing link in the empirical studies using micro level data. In this study, we use total budgetary expenditure by the local government to proxy the public spending (*pub*).[20] The variable is log-transformed so that the coefficient measures the effect of a percentage change in public spending.

In Figures (4-8), these exogenous variables are plotted on the map of administrative divisions. There is a clear spatial pattern in *ex*, *comp*, and *pub*. Evidently firms in the metro centers along the eastern coast are export-oriented. Competition, measured by HHI, is more fierce in regional centers, including those located in central and western China. The spatial variation in public spending is less pronounced, but the urban cores in the Yangtze River Delta and the Pearl River Delta receive far more public spending than the rest of the nation.

Our model allows local market conditions and institutional factor (*spec*, *comp*, and *pub*) to influence firms in neighboring districts/counties. Therefore, their spatial lags $\mathbf{W}_{2t}spec_{kt}$, $\mathbf{W}_{2t}comp_{kt}$, and $\mathbf{W}_{2t}pub_{kt}$ are also included as regressors. Finally, we end up with the following empirical model

$$\begin{aligned} tfp_{ikt} = \lambda_1 \mathbf{W}_{1t} tfp_{ikt} + \lambda_2 \mathbf{W}_{2t} tfp_{ikt} + \beta_1 rd_{kt} + \beta_2 ex_{kt} \\ + \beta_3 spec_{kt} + \beta_4 comp_{kt} + \beta_5 pub_{kt} \\ + \beta_6 \mathbf{W}_{2t} spec_{kt} + \beta_7 \mathbf{W}_{2t} comp_{kt} + \beta_8 \mathbf{W}_{2t} pub_{kt} \\ + \alpha + \text{error term}, \end{aligned} \tag{26}$$

where the error term is either $\boldsymbol{\mu} + (I - \rho \mathbf{W}_1)^{-1} \boldsymbol{\epsilon}$ in the fixed effects model or similarly $(I - \rho \mathbf{W}_1)^{-1} (\boldsymbol{\mu} + \boldsymbol{\epsilon})$ in the random effects model.

In practice, we retain the observations that have both types of neighbors, then those with complete observations. This results in an effective sample size of 84727 if the neighbor relationship among district/counties is defined by contiguity and 81331 if the neighbor relationship is defined by the 50 km criterion.

7 Empirical results

7.1 The baseline model

Table 8 summarizes the estimates of (26) in different model specifications. The conventional fixed effects estimates are reported in column (FE) as benchmark. Column (FE-IV) reports the the instrument variable 2SLS estimator discussed in Section (5.2).

[20]The expenditure categories in different issues of the *Chinese Prefecture City and County Statistical Material* are not compatible.

Since we have included two spatial lags in (26), the list of instruments (12) has to be expanded. In practice, the instruments considered are *rd*, *ex*, *spec*, *comp*, *pub*, and their spatial lags by $\mathbf{W}_1$, $\mathbf{W}_2$, or their interactions up to the second power. Since the error terms in the within transformed model are potentially correlated in space, the conventional HA and HAC type standard errors are questionable. The standard errors reported here are obtained through 50 bootstrap sample of firms. In column (FE-GMM), we implement the conventional two-step GMM using the same set of instruments. The same bootstrapping procedure is used to obtain the standard errors. Finally, we report the FG2SLS estimates of the random effects model in column (RE-FG2SLS). Note that we use $\mathbf{M} = \mathbf{W}_1$ in the error component, and the instruments suggested by (19) are built on (1) *rd*, *ex*, *spec*, *comp*, *pub*, the vector of ones, and (2) their spatial transformations by $\mathbf{W}_1$, $\mathbf{W}_2$, or their interactions up to the second power. These variables are within-transformed (except for the vector of ones) and between-transformed into the instruments. Since the error terms in the GLS transformed structural equation (17) are spherical, we report the conventional standard errors in column (RE-FG2SLS).

The estimates in column (FE-IV) are notably different from those in column (FE), indicating substantial endogeneity bias in the latter. We note that the clustered standard errors for the 2SLS within estimator are sizably smaller (not reported) than the bootstrapped values. This observation justifies our earlier concerns. According to these estimates, the productivity of a firm increases by four percent if the productivity of neighboring firms in the same district/county increases uniformly by ten percent.[21] Judged by the magnitude and significance, the spatial spillovers within the same district/county are strong. The coefficient on $\mathbf{W}2\_tfp$ is much smaller in size and insignificant. The result echoes the findings made by other researchers that spatial interactions attenuate rapidly in distance (Rosenthal and Strange, 2003, among others). Recently Baltagi, Egger, and Kesina (2015) analyze the spillover effects among Chinese firms using a statistical framework similar to ours. Interestingly, they uses the same data source as ours. There, they made an unusual observation that the strength of spillover effects, measured by the size of the spatial AR coefficient, does not change much as they extend the geographic scope from districts/counties to prefecture units, then to provinces. By incorporating two different spatial lags, our model is able to address this issue in an explicit way.[22] It is worthwhile to note that the within model sweeps off idiosyncratic effects, thus it estimates how fluctuations in productivity propagate over space. It does not reflect the selection and sorting effects suggested by the recent literature (e.g. Behrens, Duranton, and Robert-Nicoud, 2014).

The coefficients on *rd* and *ex* are both highly significant. According to these estimates, firms are more productive as they increase their development of new products or export more. These results are in line with the empirical evidence in the literature, especially those on China. The result suggests that the level of specialization in

[21] Our spatial weight matrices are row-normalized unless the firm is an island without neighbors.

[22] Their spatial model uses a single spatial AR term which corresponds to $\mathbf{W}_1$ in ours, thus is less flexible than ours.

the local area or neighboring districts/counties has little effect on firm productivity. In contrast, productivity benefits significantly from competition.[23] Both results are similar to those of Glaeser, Kallal, Scheinkman, and Shleifer (1992). Finally, we find strong evidence that firms benefit from public expenditures. A ten percent increase in local public expenditure increases productivity by 1.4 percent. The marginal effects of *comp* and *pub* in neighboring districts/counties are much smaller in size, but still highly significant. These show firms also benefit from favorable market conditions in neighboring areas.

The two-stage GMM estimates reported in column (FE-GMM) are very close to the 2SLS estimates. Despite its theoretical advantage in efficiency, the GMM procedure yields virtually identical standard errors. Therefore, we prefer the 2SLS estimator because it is much easier to implement.[24] It is noteworthy that the estimated idiosyncrasies $\hat{\boldsymbol{\mu}}$ in (7) is strongly correlated with the regressors $\mathbf{Z} = (\mathbf{Wy}, \mathbf{X})$, namely, $\text{cor}\left(\hat{\boldsymbol{\mu}}, \mathbf{Z}\hat{\boldsymbol{\delta}}_W\right) = -0.33$. Furthermore, the standard error of $\boldsymbol{\mu}$ (1.137) is large relative to that of $\mathbf{Q}_0\mathbf{u}$ (0.671). These observations invalidates the random effects assumption. The random effect FG2SLS estimates are reported in the last column. They are in sharp contrast to the 2SLS estimates, while the standard errors are notably smaller in size. The evidence thus rejects the random effects assumption.[25]

We rerun the regressions using the distance based definition of neighborhood. The estimates are summarized in Table 9. We observe a similar patter as in Table 8: the 2SLS and GMM routines produce similar estimates for the within model, which differ from the conventional within estimates or the random effects FG2SLS estimates. The difference is more pronounced in the spatial AR coefficients. Using the 2SLS estimates, we find $\text{cor}\left(\hat{\boldsymbol{\mu}}, \mathbf{Z}\hat{\boldsymbol{\delta}}_W\right) = -0.32$, and a large standard error for $\boldsymbol{\mu}$ (1.133) compared to that of $\mathbf{Q}_0\mathbf{u}$ (0.667). Thus we favor the within model over the random effects model. Again, the GMM estimator does not show a clear advantage in terms of efficiency, so we base our inference on the 2SLS estimates. The estimates based on the alternative neighborhood concept are comparable to the previous ones. The spillovers from firms in the same district/county remain significant, but slightly weaker. The spillovers from neighboring areas remain insignificant. Among the exogenous regressors, the specialization index is again insignificant, while all other regressors are highly significant with expected signs.

---

[23]The HHI is inversely related to the number of firms, which is a commonly used measure for the agglomeration. Thus, we also find evidence that agglomeration in general boosts productivity.

[24]The two-stage GMM procedure is computationally burdensome. For every bootstrap sample, the GMM criterion function has to be minimized twice. It took roughly 100 minutes to finish 50 repetitions on modern hardware.

[25]A formal Hausman test on the random effects specification demands theoretical development on the joint distribution of $\hat{\boldsymbol{\delta}}_W$ and $\hat{\boldsymbol{\delta}}_R$, which hasn't been accomplished at this stage. Given the huge difference between the two sets of estimates, the random effects specification is not likely to survive such a test.

7.2 Urban districts and distance

The baseline model shows strong and significant technological spillovers among firms in the same district/county. It also shows that the spillovers become much weaker and insignificant when the spatial linkage is extended to include firms in neighboring areas. This section further investigates the key socioeconomic or geographic factors behind these spillover effects. The first factor that comes to mind is China's administrative division. For historical reasons, urban districts and counties are very different in socioeconomic characteristics. The traditional urban districts serve as the administrative and economic centers of prefecture-level cities. They are small in area, but equipped with high quality public infrastructure. Starting from the early '90s, a new type of urban districts emerged. They are converted from counties to host the growing body of manufacturing firms. These new urban districts have experienced rapid growth in industrial output, employment, and infrastructure. Some of them have grown into new urban centers. Compared to counties or county-level cities, both types of urban districts have a higher concentration of firms and employment, but smaller geographic areas (Table 6).[26] The distinction suggests that the spillover effects studied in the baseline model may behave differently in urban districts and counties.

We divide the sample into two sub-samples by administrative type, and run the baseline regression on them. The estimates are reported in Table 10. The numbers in the first column are taken from Table 9 (column FE-IV). The second and third columns report estimates from the sub-samples. There is a sharp contrast in the estimated AR coefficients. λ_1 is highly significant in both samples, but the estimate is smaller in the urban sample. λ_2 estimated from the urban sample is much more significant and larger in size than that of the county sample. To formally test whether the AR coefficients are different between the sub-samples, we introduce a dummy variable (*county*) for counties and make it interact with $\mathbf{W}_1 tfp$ and $\mathbf{W}_2 tfp$.[27] The estimates are reported in column 4. Clearly, the difference in λ_1 is highly significant. We also perform the analysis using the contiguity based neighbor relationship (Table 11). There we observe the same pattern.

Judged by these estimates, the spatial autoregressive structure is very different in urban districts and counties. Firms in urban districts are subject to both types of spillover effects (intra-regional and inter-regional). The estimates of λ_1 and λ_2 are comparable in size because the two effects are equally potent. For firms in counties, the inter-regional spillover effect is very weak in size and significance. Therefore, the intra-regional effect plays the dominant role, and λ_1 is large in size. An easy explanation to this observation can be based on the special role of urban districts in China's administrative hierarchy. They are designated regional hubs, and have tight economic linkages with the rest of the prefecture, including their neighbors. Counties are stand-alone administrative units under the prefecture. Consequently

[26] The 2005-2007 sub-sample consists of 538 urban districts and 483 counties or county-level cities. They host 9539 and 5815 firms (annual average) in sector 3900, respectively.

[27] At the same time, the set of instruments are expanded to include their interactions with the dummy variable.

such inter-regional linkages are less close for counties.

The new economic geography theory suggests another factor that may also explain the urban-county difference: distance. Urban districts are on average much smaller in area than counties. Consequently, they are closer to their neighbors in space. Firms located in urban districts are subject to both types of spillovers, but those located in counties are hardly affected by inter-regional spillovers because of greater distance. This argument also suggests significant λ_1 and λ_2 for urban districts (smaller area); a large and significant λ_1 for counties (larger area). In order to ascertain the true mechanism behind the urban-county difference, we conduct a similar investigation into the second argument.

We sort all 2866 administrative units by area. The lower 50% are marked as small, and the rest are marked as large.[28] We then construct a dummy variable *large* to identify the large administrative units. It can be seen from Table 7 that 75% of the administrative units in the sample are small ones, hosting roughly 73% of the firms. Evidently the majority of urban districts are small, but over 50% of the counties are also small. The sample correlation between *county* and *large* is 0.269. Despite the overlap, they actually measure different concepts.

The following analysis is similar to what we have done previously. We run the baseline model on the sample of small administrative units, then on the sample of large ones. Then we let *large* interact with $\mathbf{W}_1 tfp$ and $\mathbf{W}_2 tfp$ and run the regression on the full sample. The results are reported in tables 12 and 13. The pattern is strikingly similar to that of tables 10 and 11. λ_1 is highly significant in all specifications, and significantly smaller in size for small administrative units. λ_2 is significant in the sample of small administrative units but insignificant in the other sub-sample. The difference in λ_2 is again significant. The evidence thus strongly supports the distance-based argument.

We thus find that administration type and spatial area are different factors that influence the spillover effects on districts/counties. They are related, though not identical concepts. A regression model that accounts for both factors helps to reveal the true causal effect, or at least, which factor is relatively more important. We thus include the interactions of $\mathbf{W}_1 tfp$ and $\mathbf{W}_2 tfp$ with both *county* and *large*. The estimates are reported in the last column of Tables 12 and 13. The results are mixed. In Table 12, where neighborhood is defined by the 50 km criterion, the interactions with $\mathbf{W}_1 tfp$ are significant but those interacting with $\mathbf{W}_2 tfp$ are not. It indicates that both administration type and spatial size matters for λ_1, which is smaller in size in urban districts and/or smaller administrative units. When we switch to contiguity based neighborhood, $\mathbf{W}_1 tfp$ and $\mathbf{W}_2 tfp$ interacting with *large* are significant. Surprisingly, the interaction of $\mathbf{W}_2 tfp$ and *county* is significant but the sign is hard to explain: It seems that counties benefit more from inter-regional spillovers than urban districts, controlling for spatial area. Except for this parameter, other estimates are all consistent with the previous ones. We conclude that both administration type and size jointly determines the strength of spatial productivity spillovers.

[28]The 50% quantile is 1552 square kilometers.

8 Concluding remarks

In this article, we analyze the determinants of firm-level productivity growth in China's electric apparatus industry. The geolocation information provided by the CASIF allows us to perform a joint estimation of intra-regional and inter-regional effects with a spatial autoregressive model. Because of theoretical limitations, not all spatial econometric methods can be applied to the unbalanced panel data set. We show that the Kelejian and Prucha (1998) fixed effect 2SLS estimator and the Mutl and Pfaffermayr (2011) random effect FG2SLS estimator can be modified and applied to an unbalanced panel. The empirical estimates of our baseline model reveal strong correlation between the individual effects and the exogenous regressors. The individual effects are also found to be large in size compared to the error terms.

Estimates of the baseline model shows strong spillovers among firms in the same district/county, while the spillover effects among neighboring administrative units are found to be small in size and insignificant. R&D and exports are found to contribute to higher productivity. Firms also benefit from competitiveness and public spending. However, specialization has little impact on the productivity of local firms. The market conditions in neighboring districts/counties have similar effects on productivity, but to a less extent.

The analyses on different types of administrative units reveal more information on the spillover effects. The special administrative and economic role of urban districts allows firms to interact more with peers in neighboring regions; while firms in small administrative units also have more inter-regional interactions because of shorter spatial distance. The empirical findings support both views. Although the spillover effects diminish in space, firms in urban districts or small administrative units benefit more from inter-regional spillovers because of their advantageous location.

The current research can be extended to provide more methodological rigor or empirical evidence. As we explained in Section 5.4, the Kelejian-Prucha type within estimator is in capable of addressing the autoregressive structure in the error component if the panel is unbalanced. Consequently, our fixed effects model (10) has non-spherical error terms, which make the 2SLS estimator inefficient. A GMM procedure that estimates all parameters in (7) will restore efficiency, regardless of data type. On the other hand, Mutl and Pfaffermayr's (2011) Hausman test could be extended to unbalanced panels. Such a test can be constructed from the 1st stage within estimator and the random effects FG2SLS estimator.

The CASIF provides a broad range of opportunities for empirical study. It would be interesting to see how the factors identified in the current study work on other industrial sectors. Another interesting topic is ownership structure. A large fraction of China's industrial firms are SOEs, which have gained increasing control over the market in recent years. Private firms and foreign-owned firms may differ substantially from SOEs in their capacities in generating or absorbing technological spillovers. A study on this topic could have important policy implications. Finally, the current study identifies public expenditure as a source of productivity growth, but it remains unclear how different types of public expenditure contribute to productivity growth. Further investigations is needed to cast some light on this question.

References

Abraham, F., J. Konings, and V. Slootmaekers (2010): "FDI Spillovers in the Chinese Manufacturing Sector," *Economics of Transition*, 18(1), 143–182.

Aschauer, D. A. (1989): "Is Public Expenditure Productive?," *Journal of Monetary Economics*, 23(2), 177–200.

Aw, B. Y., S. Chung, and M. J. Roberts (2000): "Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China)," *The World Bank Economic Review*, 14(1), 65–90.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2009): "R&D Investment, Exporting, and Productivity Dynamics," Discussion paper, National Bureau of Economic Research.

Baltagi, B. H., P. H. Egger, and M. Kesina (2015): "Sources of Productivity Spillovers: Panel Data Evidence from China," *Journal of Productivity Analysis*, 43(3), 389–402.

Bao, Q., J. Huang, and Y. Wang (2015): "Productivity and Firms Sales Destination: Chinese Characteristics," *Review of International Economics*.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122(3), 507–553.

Bernard, A. B., and J. B. Jensen (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47(1), 1–s25.

Boeing, P., E. Mueller, and P. G. Sandner (2015): "China's R&D Explosion–Analyzing Productivity Effects Across Ownership Types and Over Time," *ZEW-Centre for European Economic Research Discussion Paper*, (15-006).

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012): "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97(2), 339–351.

——— (2014): "Challenges of Working with the Chinese NBS Firm-Level Data," *China Economic Review*, 30(9), 339–352.

Chandra, P., and C. Long (2013): "VAT Rebates and Export Performance in China: Firm-Level Evidence," *Journal of Public Economics*, 102, 13–22.

Chen, M., and A. Guariglia (2013): "Internal Financial Constraints and Firm Productivity in China: Do Liquidity and Export Behavior Make a Difference?," *Journal of Comparative Economics*, 41(4), 1123–1140.

Clerides, S. K., S. Lach, and J. R. Tybout (1998): "Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco," *Quarterly Journal of Economics*, pp. 903–947.

Dai, M., M. Maitra, and M. Yu (2012): "Unexceptional Exporter Performance in China? The Role of Processing Trade," Working paper, China Center for Economic Research, Peking University.

De Loecker, J. (2007): "Do Exports Generate Higher Productivity? Evidence from Slovenia," *Journal of International Economics*, 73(1), 69–98.

Demurger, S. (2001): "Infrastructure Development and Economic Growth: an Explanation for Regional Disparities in China?," *Journal of Comparative Economics*, 29(1), 95–117.

Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity: Estimating Endogenous Productivity," *The Review of Economic Studies*, 80(4), 1338–1383.

Du, J., X. Liu, and Y. Zhou (2014): "State Advances and Private Retreats? Evidence of Aggregate Productivity Decomposition in China," *China Economic Review*, 31, 459 – 474.

Earle, J. S., and S. Estrin (2003): "Privatization, Competition, and Budget Constraints: Disciplining Enterprises in Russia," *Economics of Planning*, 36(1), 1–22.

Easterly, W., and R. Levine (2002): "It's Not Factor Accumulation: Stylized Facts and Growth Models," .

Elhorst, J. P. (2014): *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.

Fernald, J. G. (1999): "Roads to Prosperity? Assessing the Link between Public Capital and Productivity," *American Economic Review*, pp. 619–638.

Gao, H., and J. Van Biesebroeck (2014): "Effects of Deregulation and Vertical Unbundling on the Performance of China's Electricity Generation Sector," *Journal of Industrial Economics*, 62(1), 41–76.

Girma, S., D. Greenaway, and R. Kneller (2003): "Export Market Exit and Performance Dynamics: a Causality Analysis of Matched Firms," *Economics Letters*, 80(2), 181–187.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992): "Growth in Cities," *Journal of Political Economy*, 100(6), 1126–1152.

Gomez-Antonio, M., and B. Fingleton (2012): "Analyzing the Impact of Public Capital Stock using the NEG Wage Equation: A Spatial Panel Data Approach," *Journal of Regional Science*, 52(3), 486–502.

Griffith, R., S. Redding, and J. Van Reenen (2004): "Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries," *Review of Economics and Statistics*, 86(4), 883–895.

Griliches, Z., and J. Mairesse (1991): "R&D and Productivity Growth: Comparing Japanese and US Manufacturing Firms," in *Productivity Growth in Japan and the United States*, pp. 317–348. University of Chicago Press.

Hall, R. E., and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output per Worker Than Others?," Discussion paper, National Bureau of Economic Research.

Holtz-Eakin, D. (1992): "Public-Sector Capital and the Productivity Puzzle," Discussion paper, National Bureau of Economic Research.

Hu, A. G., and G. H. Jefferson (2004): "Returns to Research and Development in Chinese Industry: Evidence from State-Owned Enterprises in Beijing," *China Economic Review*, 15(1), 86–107.

Hu, A. G., G. H. Jefferson, and Q. Jinchang (2005): "R&D and Technology Transfer: Firm-level Evidence from Chinese Industry," *Review of Economics and Statistics*, 87(4), 780–786.

HU, C., Z. XU, AND N. YASHIRO (2015): "Agglomeration and Productivity in China: Firm Level Evidence," *China Economic Review*, 33(4), 50–66.

HU, G. (2001): "Ownership, Government R&D, Private R&D, and Productivity in Chinese Industry," *Journal of Comparative Economics*, 29(1), 136 – 157.

HULTEN, C. R., AND R. M. SCHWAB (1991): "Public Capital Formation and the Growth of Regional Manufacturing Industries," *National Tax Journal*, pp. 121–134.

JEFFERSON, G. H., R. T. G., W. LI, AND Y. ZHENG (2000): "Ownership, Productivity Change, and Financial Performance in Chinese Industry," *Journal of Comparative Economics*, 28(4), 786 – 813.

JEFFERSON, G. H., AND J. SU (2006): "Privatization and Restructuring in China: Evidence from Shareholding Ownership, 1995–2001," *Journal of Comparative Economics*, 34(1), 146–166.

JEON, Y., B. I. PARK, AND P. N. GHAURI (2013): "Foreign Direct Investment Spillover Effects in China: Are They Different across Industries with Different Technological levels?," *China Economic Review*, 26(9), 105–117.

KELEJIAN, H. H., AND I. R. PRUCHA (1998): "A Generalized Spatial Two-Stage Least Squares Procedure of Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17(1), 99–121.

KELEJIAN, H. H., I. R. PRUCHA, AND Y. YUZEFOVICH (2004): "Instrumental Variable Estimation of A Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results," in *Advances in Econometrics, Volume 18*, ed. by J. P. Lesage, and R. K. Pace, pp. 163–198. Emerald Group Publishing Limited, United Kingdom.

KELLER, W., AND S. R. YEAPLE (2009): "Multinational Enterprises, International Trade, and Productivity Growth: Firm-level Evidence from the United States," *The Review of Economics and Statistics*, 91(4), 821–831.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70(2), 317–341.

LIGTHART, J. E., AND R. M. M. SUAREZ (2011): "The Productivity of Public Capital: A Meta-Analysis," in *Infrastructure Productivity Evaluation*, pp. 5–32. Springer.

LIN, P., Z. LIU, AND Y. ZHANG (2009): "Do Chinese Domestic Firms Benefit from FDI Inflow? Evidence of Horizontal and Vertical Spillovers," *China Economic Review*, 20(4), 677–691.

LIU, Z. (2008): "Foreign Direct Investment and Technology Spillovers: Theory and Evidence," *Journal of Development Economics*, 85(1), 176–193.

LU, D. (2010): "Exceptional Exporter Performance? Evidence from Chinese Manufacturing Firms," Working paper, University of Chicago.

LU, J., Y. LU, AND Z. TAO (2010): "Exporting Behavior of Foreign Affiliates: Theory and Evidence," *Journal of International Economics*, 81(2), 197–205.

LUO, X., AND N. ZHU (2015): "What Drives Productivity Volatility of Chinese Industrial Firms?," Discussion paper, CIRANO.

MARSHALL, A. (1890): *Principles of Economics*. Macmillan, London.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Melo, P. C., D. J. Graham, and R. Brage-Ardao (2013): "The Productivity of Transport Infrastructure Investment: A Meta-Analysis of Empirical Evidence," *Regional Science and Urban Economics*, 43(5), 695–706.

Munnell, A. H. (1992): "Policy Watch: Infrastructure Investment and Economic Growth," *The Journal of Economic Perspectives*, pp. 189–198.

Mutl, J., and M. Pfaffermayr (2011): "The Hausman Test in a Cliff and Ord Panel Model," *The Econometrics Journal*, 14(1), 48–76.

Nickell, S. J. (1996): "Competition and Corporate Performance," *Journal of Political Economy*, pp. 724–746.

Nie, H., T. Jiang, and R. Yang (2012): "Chinese Industrial Enterprise Database: Current Use and Potential Problems," *Journal of World Economy*, 143(5), 142–158.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Pereira, A. M., and J. M. Andraz (2003): "On the Impact of Public Investment on the Performance of US Industries," *Public Finance Review*, 31(1), 66–90.

Pisu, M. (2008): "Export Destinations and Learning-by-Exporting: Evidence from Belgium," *National Bank of Belgium Working Paper*, (140).

Porter, M. E. (1990): *The Competitive Advantage of Nations*. Free Press, New York.

Rosenthal, S. S., and W. C. Strange (2003): "Geography, Industrial Organization, and Agglomeration," *Review of Economics and Statistics*, 85(2), 377–393.

Rosenthal, S. S., and W. C. Strange (2004): "Evidence on the Nature and Sources of Agglomeration Economies," *Handbook of regional and urban economics*, 4, 2119–2171.

Schmerer, H.-J., and L. Wang (2014): "Firm Performance and Trade with Low-Income Countries: Evidence from China," Discussion Paper CESIFO Working Paper No.4934, Institute for Employment Research, Germany.

Schmitz Jr, J. A. (2005): "What Determines Productivity? Lessons from the Dramatic Recovery of the US and Canadian Iron Ore Industries Following their Early 1980s Crisis," *Journal of Political Economy*, 113(3), 582–625.

Sheng, Y., and L. Song (2013): "Re-estimation of Firms' Total Factor Productivity in China's Iron and Steel Industry," *China Economic Review*, 24(3), 177–188.

Silva, A., O. Afonso, and A. P. Africano (2012): "Which Manufacturing Firms Learn by Exporting?," *The Journal of International Trade and Economic Development*, 21(6), 773–805.

Song, L., and Y. Yao (2004): "Impacts of Privatization on Firm Performance in China," *China Center for Economic Research Working Paper Series (July 28, 2004)*.

Sun, X., and J. Hong (2011): "Exports, Ownership and Firm productivity: Evidence from China," *The World Economy*, 34(7), 1199–1215.

Syverson, C. (2004): "Market Structure and Productivity: A Concrete Example," Discussion paper, National Bureau of Economic Research.

——— (2010): “What Determines Productivity?,” Discussion paper, National Bureau of Economic Research.

Tian, W., and M. Yu (2014): “Outward Direct Investment and Firm Productivity: Evidence from Chinese Firms,” Working paper, China Center for Economic Research, Peking University.

Vijverberg, W. P., F.-C. Fu, and C.-P. C. Vijverberg (2011): “Public Infrastructure as a Determinant of Productive Performance in China,” *Journal of Productivity Analysis*, 36(1), 91–111.

Wagner, J. (2012): “International Trade and Firm Performance: a Survey of Empirical Studies since 2006,” *Review of World Economics*, 148(2), 235–267.

Wei, Y., and X. Liu (2006): “Productivity Spillovers from R&D, Exports and FDI in China’s Manufacturing Sector,” *Journal of International Business Studies*, 37(4), 544–557.

Wilhelmsson, F., and K. Kozlov (2007): “Exports and Productivity of Russian firms: in Search of Causality,” *Economic Change and Restructuring*, 40(4), 361–385.

Yu, N., M. De Jong, S. Storm, and J. Mi (2013): “Spatial Spillover Effects of Transport Infrastructure: Evidence from Chinese Regions,” *Journal of Transport Geography*, 28, 56–66.

Yueh, L. (2010): “Accounting for China’s Labour Productivity Growth in the 2000s,” Discussion Paper CGC Discussion Paper No. 1, China Growth Centre (CGC), University of Oxford.

Table 1: Two-digit industrial sectors, GB/T 4754-2002

industry code	industry name	industry code	industry name
600	coal mining and washing	2700	pharmaceutical manufacturing
700	oil and natural gas extraction	2800	synthetic fabric manufacturing
800	ferrous metal extraction and washing	2900	rubber manufacturing
900	nonferrous metal extraction and washing	3000	plastic manufacturing
1000	non-metallic mineral extraction and washing	3100	non-metallic mineral processing
1100	other mining	3200	ferrous metal smelting and rolling
1300	agricultural products processing	3300	nonferrous metal smelting and rolling
1400	food products manufacturing	3400	metal products manufacturing
1500	beverage products manufacturing	3500	general-purpose equipment manufacturing
1600	tobacco products manufacturing	3600	special equipment manufacturing
1700	miscellaneous synthetic fabric products manufacturing	3700	transport equipment manufacturing
1800	garments, shoes, and hats manufacturing	3900	electric apparatus manufacturing
1900	leather and leather products manufacturing	4000	manufacturing of communication equipments, computers, and other electronic equipments
2000	wood processing and wood products manufacturing	4100	meters and office machinery manufacturing
2100	furniture manufacturing	4200	art ware and other manufacturing
2200	paper and paper products manufacturing	4300	waste products recycling and processing
2300	printed matter and audiovisual products production	4400	production and supply of electric power and heat
2400	stationary and sports products manufacturing	4500	production and supply of natural gas
2500	petroleum processing, coking and nuclear fuel processing	4600	production and supply of running water
2600	chemical material and chemical products manufacturing		

Table 2: Summary Statistics of CASIF, 1996-2010

year	full sample			above-scale firms[1]			percent of national aggregate[2]			change in sample[3]	
	count	output	employees	count	output	employees	count	output	employees	in	out
1996	23480	3383.75	375.16	21921	3349.19	365.76	–	–	–	–	–
1997	23889	3653.88	369.57	21671	3578.76	356.59	–	–	–	2374	1965
1998	163810	6690.22	612.80	123351	6525.97	563.84	74.7%	96.3%	59.0%	143162	3241
1999	160623	7174.11	573.57	120709	7001.87	531.34	74.5%	96.3%	91.5%	23801	26988
2000	161401	8441.10	549.41	125140	8274.83	512.93	76.8%	96.6%	92.2%	26088	25310
2001	159703	8772.10	503.62	129458	8616.13	473.15	75.6%	90.3%	87.0%	39519	41217
2002	179871	10915.00	545.11	150171	10757.41	517.95	82.7%	97.1%	93.8%	41862	21694
2003	194452	14018.17	567.64	170950	13886.01	547.60	87.1%	97.6%	95.3%	41265	26684
2004	276826	15636.75	654.10	259219	15627.63	645.00	93.8%	77.5%	97.4%	126147	43773
2005	269752	24834.29	681.88	256486	24652.75	672.47	94.4%	98.0%	97.5%	38885	45959
2006	299599	31198.10	727.40	287773	30906.18	720.55	95.3%	97.6%	97.9%	55030	25183
2007	334151	39943.59	778.38	327118	39730.99	774.30	97.1%	98.1%	98.3%	61981	27429
2008	260106	37652.23	639.00	257080	37548.71	637.28	60.3%	74.0%	72.1%	7787	81832
2009	202724	32049.91	505.83	200863	32032.94	505.06	46.2%	58.4%	57.2%	0	57382
2010	342077	85057.47	1290.07	335441	84824.55	1286.28	74.1%	121.4%	134.76%	160572	21219

[1]Above-scale firms are those with an annual revenue from principal business over 5 million CNY.
[2]The sub-total of above-scale firms in the sample as a percentage of the national aggregate reported by the *China Statistical Yearbook* (2012 issue).
[3]Numbers of entrants and exists in the current year.

Table 3: CASIF key variables and availability, 1996-2010

Variable name	Available in	Number of missing obs (10,000 or more)
organization code	all years	10,000 in 2001, 161,643 in 2008, 121,232 in 2009
legal name	all years	10,000 in 2001, 37,940 in 2010
province	all but 1997-2000	10,000 in 2001, 17,516 in 2002
prefecture	all but 1996-2000	36,474 in 2001, 30,309 in 2002, 25,003 in 2003, 41,227 in 2004, 23,919 in 2005, 23,670 in 2006, 23,033 in 2007, 25,099 in 2008, 16,517 in 2009
urban district or county	all but 1996-2000	44,849 in 2001, 42,660 in 2002, 72,922 in 2004
township	all but 2003	13,720 in 1996 and 1997, over 40,000 in 1998-2002, about 100,000 in 2004-2009
street address	all but 2003 and 2010	over 13,000 in 1996 and 1997, about 90,000 in 1998-2002, about 170,000 in 2004-2009
administrative code	all years	10,000 in 2001
industry classification code	all years	10,000 in 2001
affiliation type	all years	10,000 in 2001
registration code	all but 1996 and 1997	10,000 in 2001
starting year	all years	10,602 in 2001
operation status	all years	12,624 in 2001
gross output (current price)	all but 2004	10,000 in 2001
gross output (constant price)	1996-2003	10,000 in 2001
output of new merchandise	all but 1996, 1997, 2004	10,000 in 2001
gross sales revenue (current price)	all but 2004	10,000 in 2001
export sales	all but 1996, 1997, 2004	10,000 in 2001
total value added	all but 2001, 2002, 2004, 2008, 2009	
total current assets	all years	10,000 in 2001
short-term investment	2001-2002, 2004-2007, 2010	10,000 in 2001
accounts receivable	all but 1996-1997	10,000 in 2001
inventory	all years	10,000 in 2001
finished goods in stock	all years	10,000 in 2001
average current asset	all but 2008-2010	10,000 in 2001
long-term investment	all but 1996, 2008, 2009	10,000 in 2001
original value of fixed assets	all but 2008 and 2009	10,000 in 2001
net value of fixed assets	all but 2008-2010	10,000 in 2001

Table3—continued

Variable name	Available in	Number of missing obs (10,000 or more)
total fixed assets	all years	10,000 in 2001
accumulated depreciation	all but 2008 and 2009	10,000 in 2001
current depreciation	all but 2008-2010	10,000 in 2001
intangible assets	all but 2008 and 2009	10,000 in 2001
total assets	all years	10,000 in 2001
current liabilities	all years	10,000 in 2001
accounts payable	2004-2010	
long-term liabilities	all years	10,000 in 2001
total liabilities	all years	10,000 in 2001
owner's equity	all years	10,000 in 2001
paid-in capital	all years	10,000 in 2001
state capital	all but 2008	10,000 in 2001
collective capital	all but 2008	10,000 in 2001
legal person's capital	all but 2008	10,000 in 2001
individual capital	all but 2008	10,000 in 2001
HMT capital	all but 2008	10,000 in 2001
foreign capital	all but 2008	10,000 in 2001
revenue from principal business	all years	10,000 in 2001
expenses of principal business	all years	10,000 in 2001
revenue from other operations	2004-2010	
profit from other operations	all years	10,000 in 2001
operating expenses	all years	10,000 in 2001
general and administrative expenses	all years	10,000 in 2001
financial expenses	all years	10,000 in 2001
operating profit	all years	10,000 in 2001
investment return	all but 1996-2000, 2003	10,000 in 2001
subsidy income	all but 2008-2010	10,000 in 2001
total profit	all years	10,000 in 2001
corporate income tax payable	all years	10,000 in 2001
business tax payable	all years	10,000 in 2001
VAT payable	all years	10,000 in 2001
research and development expenses	2001-2002, and 2004-2007	10,000 in 2001

Table3—continued

Variable name	Available in	Number of missing obs (10,000 or more)
interest expenses	all years	10,000 in 2001
total wage payable	all but 2008-2010	10,000 in 2001
wage payable for principal business	all but 2008-2010	10,000 in 2001
total welfare expenses payable	all but 2008-2010	10,000 in 2001
welfare expenses payable for principal business	all but 2008-2010	10,000 in 2001
total intermediate input	all but 2008-2010	10,000 in 2001

Table 4: Operation status

Code	Status
1	set up
2	shut down
3	under construction
4	withdraw
9	others

Table 5: Affliation type

Code	Affiliation Type
10	central government
20	provincial government
40	prefecture government
50	county government
61	neighborhood
62	township
63	village
71	urban community
72	village community
90	others

Table 6: Summary statistics of urban districts and counties, sample average in 2005-2007

	urban districts					counties				
	min	1st quartile	median	3rd quartile	max	min	1st quartile	median	3rd quartile	max
area[1]	7.3	105.4	324.8	810.0	6653.0	167.5	1004.0	1449.0	2103.0	8928.0
number of firms	0.7	3.0	6.0	15.3	755.3	0.7	2.0	3.7	9.3	466.7
employment	10.0	369.2	1160.0	3472.0	357200.0	17.0	178.3	497.0	1682.0	90050.0
public expenses[1]	70.3	299.9	577.6	1016.0	63280.0	164.4	488.4	677.8	975.4	6064.0

[1] Units: square kilometer (area) and RMB 1 million (public expenditure).

Table 7: Sample division by administration type and size, 2005-2007 average

	small		large	
	number of admin units	number of firms	number of admin units	number of firms
urban districts	492	7766.3	46	1772.7
counties	270	3369.7	213	2445.3

Table 8: The baseline model: contiguous districts/counties treated as neighbors, 84727 observations on 26174 distinct firms

Dependent variable: *tfp*

regressor	FE[1]	FE-IV[2]	FE-GMM[2]	RE-FG2SLS
W1_tfp	0.185**	0.403**	0.410**	0.673**
	(0.011)	(0.074)	(0.075)	(0.008)
W2_tfp	0.130**	0.111	0.118	0.134**
	(0.013)	(0.096)	(0.092)	(0.008)
rd	0.181**	0.192**	0.185**	0.389**
	(0.027)	(0.030)	(0.029)	(0.016)
ex	0.083**	0.088**	0.087**	0.070**
	(0.025)	(0.028)	(0.028)	(0.009)
spec	0.019*	-0.006	-0.006	-0.011**
	(0.009)	(0.014)	(0.014)	(0.002)
comp	-0.634**	-0.832**	-0.863**	0.132**
	(0.075)	(0.123)	(0.123)	(0.015)
pub	0.234**	0.137**	0.132**	-0.009**
	(0.013)	(0.036)	(0.036)	(0.003)
W2_spec	-0.030**	-0.021	-0.019	-0.035**
	(0.009)	(0.013)	(0.013)	(0.002)
W2_comp	-0.573**	-0.421**	-0.400**	0.031*
	(0.077)	(0.135)	(0.133)	(0.017)
W2_pub	0.030**	0.019**	0.018**	-0.002
	(0.005)	(0.006)	(0.006)	(0.002)
Intercept	1.477**	1.279**	1.234**	1.261**
	(0.139)	(0.159)	(0.158)	(0.040)
R^2	0.13	0.12	—	0.73
$\hat{\sigma}_\mu$	1.166	1.137	—	1.399
$\hat{\sigma}_u$ $(\hat{\sigma}_\epsilon)^3$	0.666	0.671	—	0.560
$\hat{\rho}$	—	—	—	-0.962

Significance codes: '**' 0.05, '*' 0.10

[1]Clustered standard errors in parentheses

[2]Bootstrapped standard errors in parentheses

[3]In columns (FE) and (FE-IV), the numbers are standard errors of $\mathbf{Q}_0\mathbf{u}$ in (10); in column (RE-FG2SLS), the number is $\hat{\sigma}_\epsilon$ estimated by (16).

Table 9: The baseline model: districts/counties within 50 kilometers treated as neighbors, 81331 observations on 24903 distinct firms

Dependent variable: *tfp*

regressor	FE[1]	FE-IV[2]	FE-GMM[2]	RE-FG2SLS
W1_tfp	0.163**	0.354**	0.360**	0.592**
	(0.011)	(0.066)	(0.070)	(0.010)
W2_tfp	0.181**	0.088	0.087	0.188**
	(0.015)	(0.081)	(0.078)	(0.009)
rd	0.179**	0.186**	0.178**	0.465**
	(0.027)	(0.027)	(0.027)	(0.017)
ex	0.083**	0.087**	0.086**	0.055**
	(0.025)	(0.026)	(0.026)	(0.009)
spec	0.020**	-0.004	-0.004	-0.002
	(0.009)	(0.011)	(0.011)	(0.002)
comp	-0.590**	-0.776**	-0.808**	0.118**
	(0.074)	(0.104)	(0.111)	(0.017)
pub	0.210**	0.168**	0.168**	-0.020**
	(0.013)	(0.036)	(0.034)	(0.003)
W2_spec	-0.042**	-0.022	-0.019	-0.050**
	(0.009)	(0.016)	(0.016)	(0.002)
W2_comp	-0.743**	-0.524**	-0.495**	0.055**
	(0.081)	(0.122)	(0.127)	(0.020)
W2_pub	0.023**	0.019**	0.019**	-0.003
	(0.004)	(0.007)	(0.006)	(0.002)
Intercept	1.689**	1.498**	1.468**	1.524**
	(0.141)	(0.191)	(0.198)	(0.042)
R^2	0.14	0.13	—	0.70
$\hat{\sigma}_\mu$	1.140	1.133	—	1.325
$\hat{\sigma}_u$ $(\hat{\sigma}_\epsilon)^3$	0.663	0.667	—	0.588
$\hat{\rho}$	—	—	—	-0.824

Significance codes: '**' 0.05, '*' 0.10

[1]Clustered standard errors in parentheses

[2]Bootstrapped standard errors in parentheses

[3]In columns (FE) and (FE-IV), the numbers are standard errors of $\mathbf{Q}_0\mathbf{u}$ in (10); in column (RE-FG2SLS), the number is $\hat{\sigma}_\epsilon$ estimated by (16).

Table 10: Spillover effects on urban districts and counties: districts/counties within 50 kilometers treated as neighbors.

Dependent variable: *tfp*[1]

regressor	baseline	urban districts	counties	full model
W1_tfp	0.354**	0.204**	0.587**	0.246**
	(0.066)	(0.079)	(0.097)	(0.065)
W2_tfp	0.088	0.203**	0.092	0.135**
	(0.081)	(0.101)	(0.099)	(0.068)
interactions:				
W1_tfp×*county*	—	—	—	0.343**
				(0.066)
W2_tfp×*county*	—	—	—	-0.088
				0.057
rd	0.186**	0.157**	0.206**	0.176**
	(0.027)	(0.039)	(0.037)	(0.028)
ex	0.087**	0.053*	0.154**	0.084**
	(0.026)	(0.031)	(0.040)	(0.025)
spec	-0.004	0.000	0.004	-0.002
	(0.011)	(0.018)	(0.015)	(0.010)
comp	-0.776**	-0.511**	-0.987**	-0.733**
	(0.104)	(0.126)	(0.152)	(0.101)
pub	0.168**	0.136**	0.162**	0.157**
	(0.036)	(0.039)	(0.034)	(0.030)
W2_spec	-0.022	-0.023	-0.035*	-0.020
	(0.016)	(0.017)	(0.019)	(0.014)
W2_comp	-0.524**	-0.480**	-0.720**	-0.503**
	(0.122)	(0.134)	(0.217)	(0.107)
W2_pub	0.019**	0.026**	-0.021**	0.021**
	(0.007)	(0.006)	(0.010)	(0.006)
Intercept	1.498**	1.979**	0.316	1.358**
	(0.191)	(0.199)	(0.315)	(0.179)
N	81331	51480	29851	81331
n	24903	15575	9328	24903
R^2	0.13	0.10	0.18	0.13
$\hat{\sigma}_\mu$	1.133	1.144	1.114	1.404
$\hat{\sigma}_u$	0.667	0.699	0.605	0.667

Significance codes: '**' 0.05, '*' 0.10

[1]All regressions estimated by fixed effects 2SLS, bootstrapped standard errors in parentheses.

Table 11: Spillover effects on urban districts and counties: contiguous units treated as neighbors.

Dependent variable: *tfp*[1]

regressor	baseline	urban districts	counties	full model
W1_tfp	0.403**	0.320**	0.542**	0.314**
	(0.074)	(0.071)	(0.112)	(0.071)
W2_tfp	0.111	0.137*	0.175	0.135
	(0.096)	(0.082)	(0.140)	(0.084)
interactions:				
$W1\_tfp \times county$	—	—	—	0.300**
				(0.067)
$W2\_tfp \times county$	—	—	—	-0.037
				(0.070)
rd	0.192**	0.167**	0.213**	0.183**
	(0.030)	(0.030)	(0.044)	(0.029)
ex	0.088**	0.057**	0.149**	0.085**
	(0.028)	(0.026)	(0.029)	(0.027)
spec	-0.006	-0.016	0.014	-0.004
	(0.014)	(0.015)	(0.019)	(0.014)
comp	-0.832**	-0.643**	-0.944**	-0.791**
	(0.123)	(0.099)	(0.153)	(0.117)
pub	0.137**	0.121**	0.123	0.124**
	(0.036)	(0.036)	(0.088)	(0.032)
W2_spec	-0.021	-0.014	-0.046**	-0.023**
	(0.013)	(0.014)	(0.022)	(0.012)
W2_comp	-0.421**	-0.359**	-0.588**	-0.405**
	(0.135)	(0.123)	(0.198)	(0.123)
W2_pub	0.019**	0.026**	-0.012	0.023**
	(0.006)	(0.008)	(0.012)	(0.006)
Intercept	1.279**	1.793**	0.045	1.168**
	(0.159)	(0.198)	(0.257)	(0.153)
N	84727	54513	30214	84727
n	26174	16679	9495	26174
R^2	0.12	0.12	0.19	0.12
$\hat{\sigma}_\mu$	1.137	1.147	1.100	1.398
$\hat{\sigma}_u$	0.671	0.705	0.603	0.672

Significance codes: '**' 0.05, '*' 0.10

[1] All regressions estimated by fixed effects 2SLS, bootstrapped standard errors in parentheses.

Table 12: Spillover effects on small and large administrative units: units within 50 kilometers treated as neighbors.

Dependent variable: *tfp*[1]

regressor	baseline	small admins	large admins	model 1	model 2
W1_tfp	0.354**	0.286**	0.584**	0.277**	0.219**
	(0.066)	(0.087)	(0.104)	(0.056)	(0.058)
W2_tfp	0.088	0.196**	-0.065	0.141*	0.117*
	(0.081)	(0.096)	(0.087)	(0.080)	(0.070)
interactions:					
W1_tfp×*large*	—	—	—	0.400**	0.317**
				(0.070)	(0.083)
W1_tfp×*county*	—	—	—	—	0.163**
					(0.075)
W2_tfp×*large*	—	—	—	-0.162**	-0.137
				(0.073)	(0.091)
W2_tfp×*county*	—	—	—	—	0.035
					(0.073)
rd	0.186**	0.206**	0.099**	0.176**	0.170**
	(0.027)	(0.029)	(0.038)	(0.027)	(0.028)
ex	0.087**	0.079*	0.116**	0.092**	0.088**
	(0.026)	(0.033)	(0.040)	(0.025)	(0.025)
spec	-0.004	-0.020	0.004	-0.016*	-0.009
	(0.011)	(0.013)	(0.023)	(0.009)	(0.010)
comp	-0.776**	-0.725**	-0.869**	-0.759**	-0.709**
	(0.104)	(0.133)	(0.190)	(0.090)	(0.091)
pub	0.168**	0.121**	0.209**	0.147**	0.164**
	(0.036)	(0.042)	(0.062)	(0.038)	(0.032)
W2_spec	-0.022	-0.018	-0.050**	-0.024*	-0.025*
	(0.016)	(0.016)	(0.020)	(0.014)	(0.013)
W2_comp	-0.524**	-0.574**	-0.403*	-0.511**	-0.526**
	(0.122)	(0.131)	(0.222)	(0.098)	(0.098)
W2_pub	0.019**	0.004	0.023**	0.012*	0.017**
	(0.007)	(0.008)	(0.009)	(0.006)	(0.006)
Intercept	1.498**	1.923**	0.474*	1.577**	1.493**
	(0.191)	(0.180)	(0.282)	(0.176)	(0.166)
N	81331	60463	20868	81331	81331
n	24903	18358	6545	24903	24903
R^2	0.13	0.11	0.20	0.13	0.13
$\hat{\sigma}_\mu$	1.133	1.118	1.136	1.384	1.507
$\hat{\sigma}_u$	0.667	0.681	0.619	0.667	0.666

Significance codes: '**' 0.05, '*' 0.10

[1] All regressions estimated by fixed effects 2SLS, bootstrapped standard errors in parentheses.

Table 13: Spillover effects on small and large administrative units: contiguous units treated as neighbors.

Dependent variable: *tfp*[1]

regressor	baseline	small admins	large admins	model 1	model 2
W1_tfp	0.403**	0.278**	0.679**	0.298**	0.278**
	(0.074)	(0.083)	(0.161)	(0.075)	(0.071)
W2_tfp	0.111	0.261**	-0.023	0.218**	0.178**
	(0.096)	(0.096)	(0.108)	(0.098)	(0.084)
interactions:					
W1_tfp×large	—	—	—	0.438**	0.520**
				(0.077)	(0.086)
W1_tfp×county	—	—	—	—	-0.034
					(0.078)
W2_tfp×large	—	—	—	-0.227**	-0.373**
				(0.085)	(0.098)
W2_tfp×county	—	—	—	—	0.247**
					(0.085)
rd	0.192**	0.202**	0.142**	0.184**	0.179**
	(0.030)	(0.036)	(0.034)	(0.030)	(0.030)
ex	0.088**	0.075**	0.122**	0.093**	0.093**
	(0.028)	(0.034)	(0.037)	(0.027)	(0.027)
spec	-0.006	-0.013	0.001	-0.015	-0.007
	(0.014)	(0.014)	(0.032)	(0.013)	(0.014)
comp	-0.832**	-0.707**	-1.006**	-0.796**	-0.754**
	(0.123)	(0.127)	(0.196)	(0.120)	(0.116)
pub	0.137**	0.097**	0.127	0.103**	0.108**
	(0.036)	(0.036)	(0.082)	(0.033)	(0.028)
W2_spec	-0.021	-0.026*	-0.042**	-0.029**	-0.032**
	(0.013)	(0.014)	(0.021)	(0.012)	(0.012)
W2_comp	-0.421**	-0.481**	-0.418**	-0.436**	-0.447**
	(0.135)	(0.118)	(0.203)	(0.124)	(0.115)
W2_pub	0.019**	0.008	0.020	0.011*	0.014**
	(0.006)	(0.007)	(0.013)	(0.006)	(0.006)
Intercept	1.279**	1.692**	0.385*	1.316**	1.258**
	(0.159)	(0.168)	(0.233)	(0.142)	(0.146)
N	84727	60218	24509	84727	84727
n	26174	18324	7850	26174	26174
R^2	0.12	0.10	0.17	0.12	0.12
$\hat{\sigma}_\mu$	1.137	1.114	1.172	1.373	1.465
$\hat{\sigma}_u$	0.671	0.683	0.638	0.671	0.672

Significance codes: '**' 0.05, '*' 0.10

[1] All regressions estimated by fixed effects 2SLS, bootstrapped standard errors in parentheses.

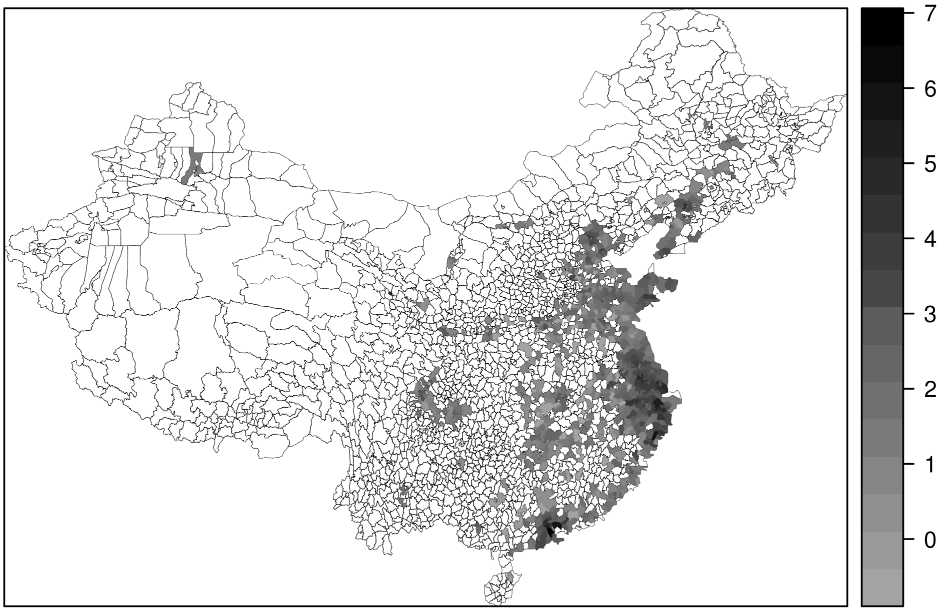

Figure 1: District/county subtotal: number of firms in sector 3900, 2005-2007 average.

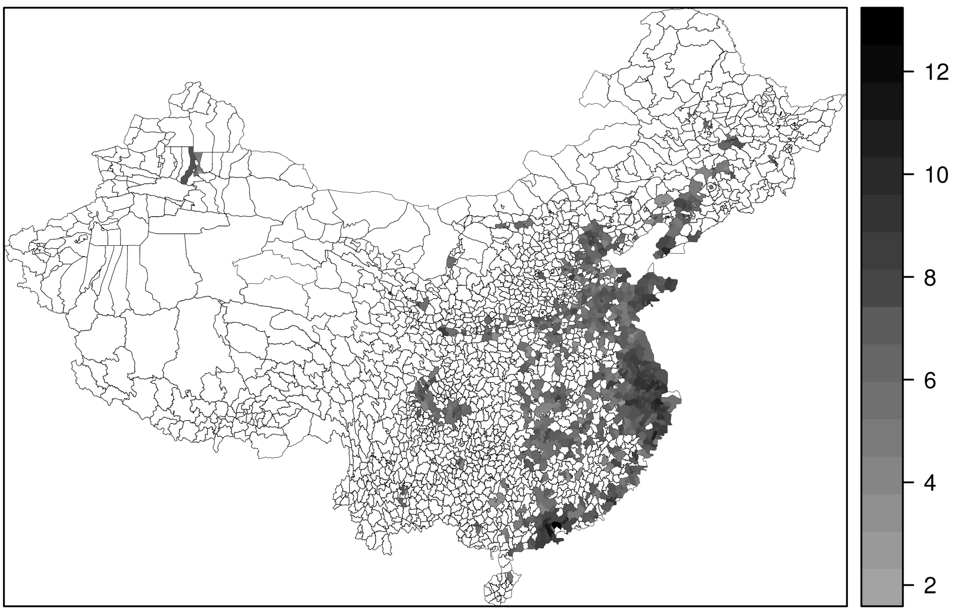

Figure 2: District/county subtotal: employment in sector 3900, 2005-2007 average.

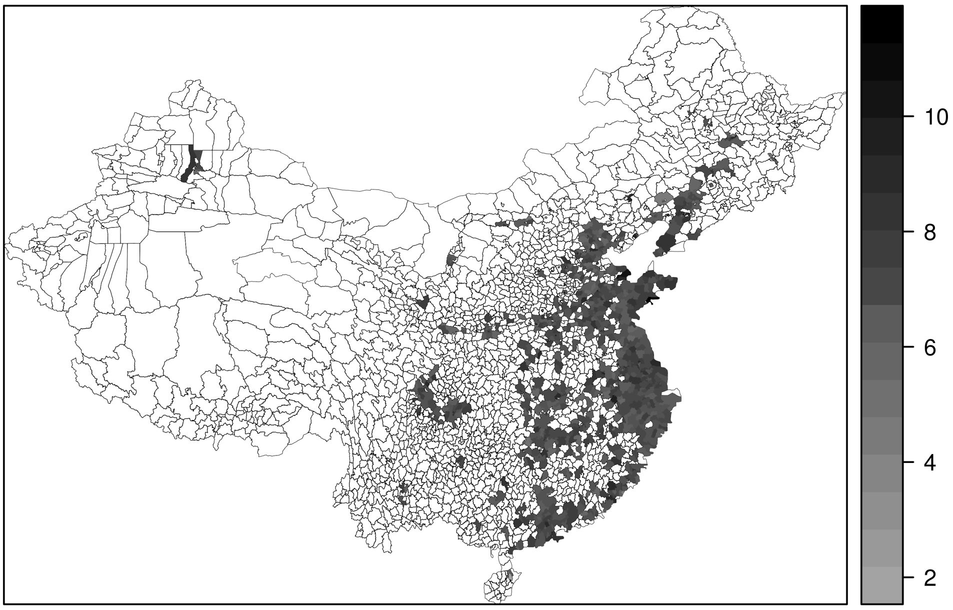

Figure 3: District/county average TFP in sector 3900, futher averaged over 2005-2007.

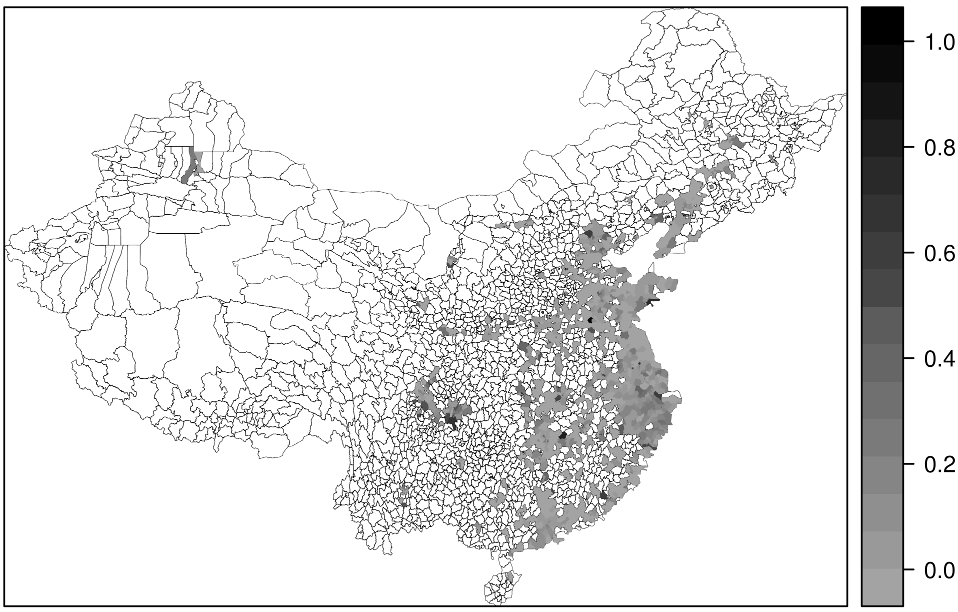

Figure 4: District/county average new product-total ouput ratio in sector 3900, futher averaged over 2005-2007.

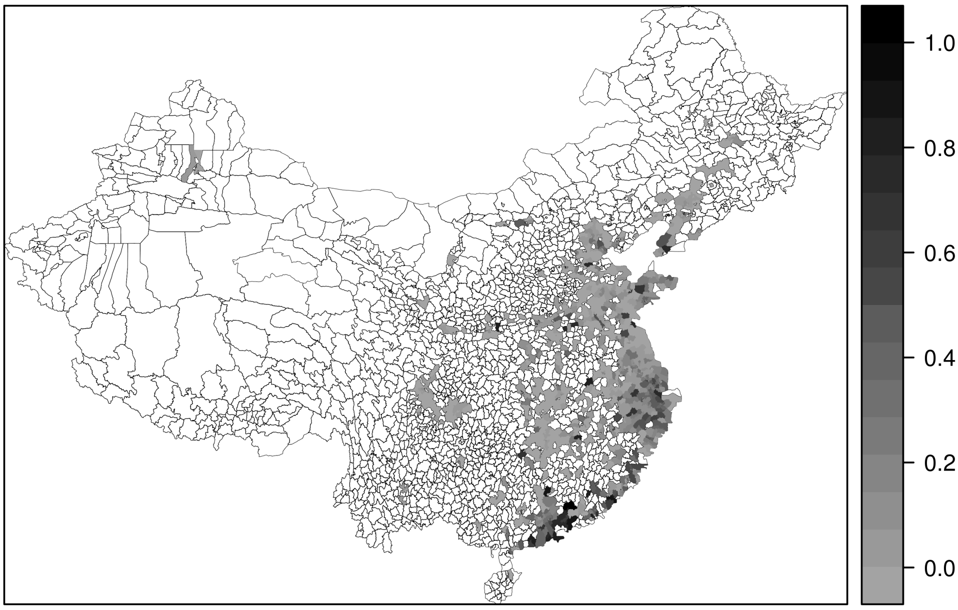

Figure 5: District/county average export-sales ratio in sector 3900, futher averaged over 2005-2007.

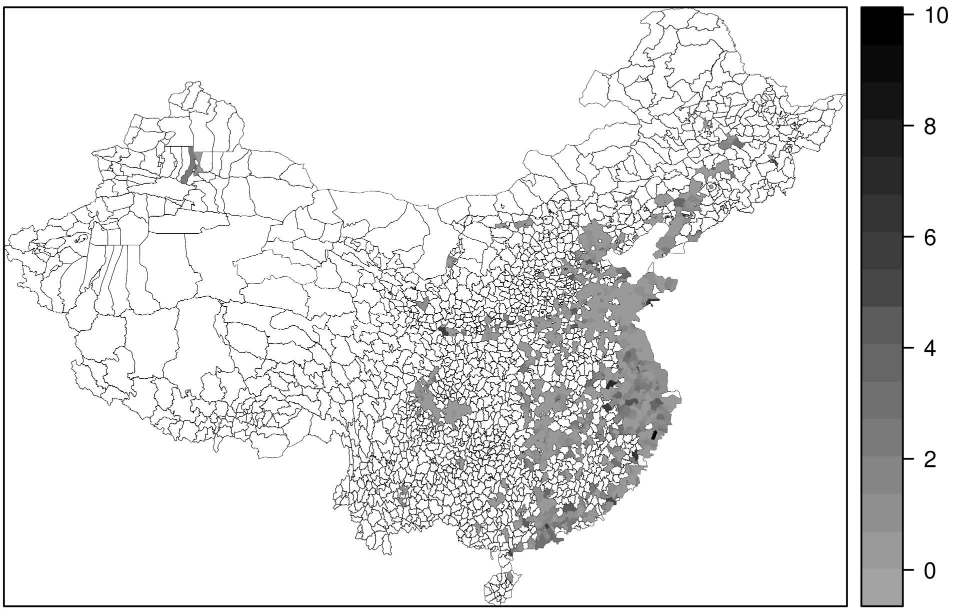

Figure 6: District/county specialization index in sector 3900, futher averaged over 2005-2007.

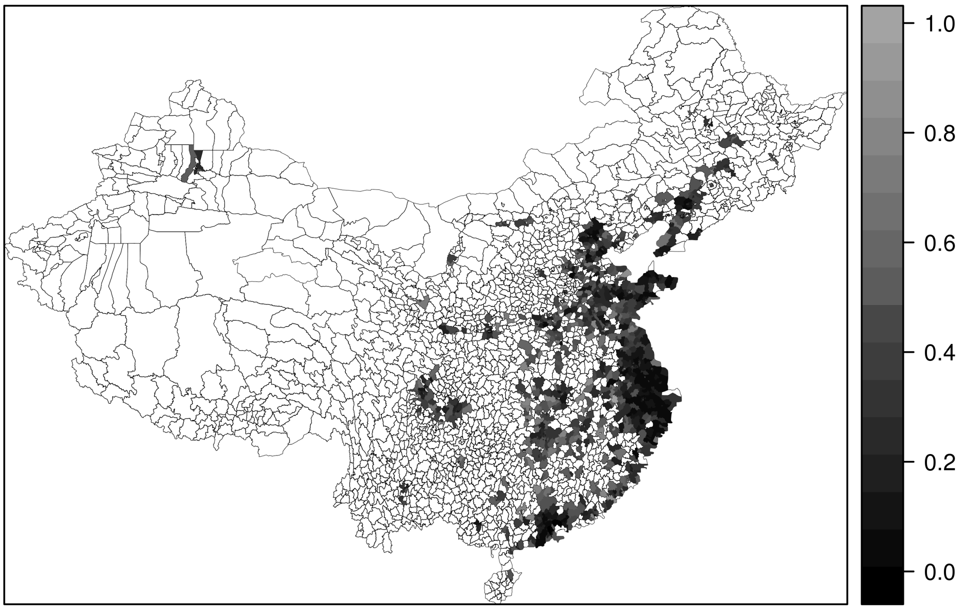

Figure 7: District/county Herfindahl-Hirschman index in sector 3900, futher averaged over 2005-2007.

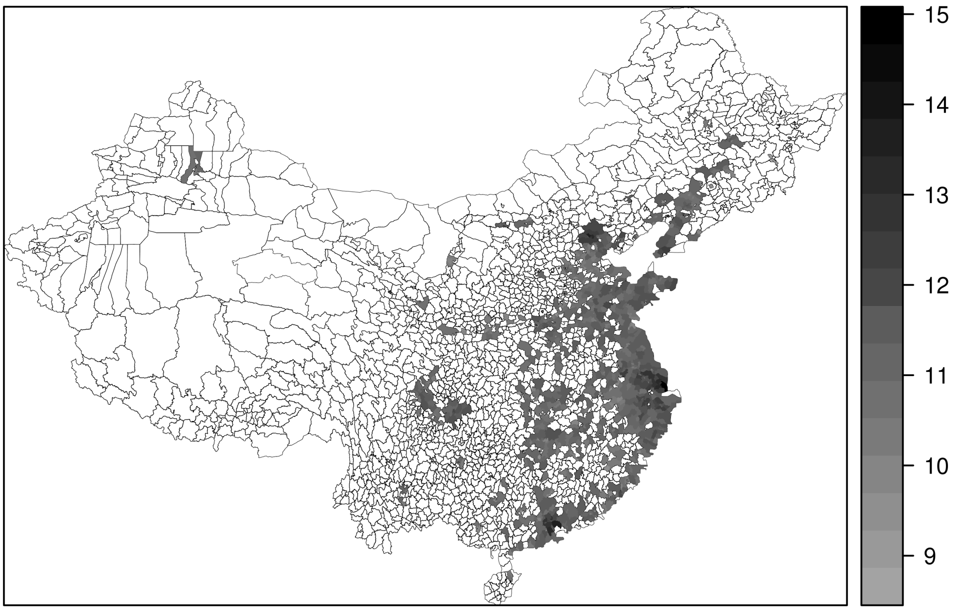

Figure 8: District/county total budgetary expenditure (log-transformed) in sector 3900, futher averaged over 2005-2007.