JCER Working Paper
AEPR series
No. 2018-1-5

This is the pre-peer- reviewed version of the following article: "Geographic Concentration in Indian Manufacturing and Service Industries: Evidence from 1998 to 2013", *Asian Economic Policy Review*, vol. 14, issue 1, which has been published in final form at https://onlinelibrary.wiley.com/doi/10.1111/aepr.12251 and DOI: 10.1111/aepr.12251.

The Spatial Development of Manufacturing and Services: Evidence from India 1998 - 2013

Amrit Amirapu
(University of Kent)
Rana Hasan
(Asian Development Bank)
Yi Jiang
(Asian Development Bank)
Alex Klein
(University of Kent)

This paper was prepared for the Twenty-seventh Asian Economic Policy Review (AEPR) Conference "India," April 7, 2018, Tokyo.

January 2019

Asian Economic Policy Review
Japan Center for Economic Research

To authors

If you want to introduce the same working paper you wrote and presented at the AEPR conference held in Tokyo on Apr. 7, 2018, in your own/your affiliation's website, please be aware the following requirements.

To ensure that all citations and references to your published article are captured by the SSCI (Social Sciences Citation Index), authors are required to amend the cover page of your working paper as soon as practical after publication in AEPR. The amended cover page should include the full article citation, journal name, volume and issue, and DOI, as well as a hyperlink to the published article. The cover page of JCER Working Paper AEPR series has been already amended after publication in AEPR. The face of this working paper is an example of an amended working paper cover page.

The Spatial Development of Manufacturing and Services: Evidence from India 1998 - 2013

Amrit Amirapu[1] Rana Hasan[2] Yi Jiang[3] Alex Klein[4]

March 2018

(Preliminary Draft for AEPR Conference)

1 Introduction

The spatial distribution of economic activity and how it evolves are issues of considerable policy importance. Where and why firms locate are questions whose answers have profound implications for both economic growth and its distribution within a country. In this paper, we use comprehensive new data to examine aspects of the spatial development of both manufacturing and services in India from 1998 to 2013, a period over which India's economic reforms of the early 1990s had solidified and had a chance to influence the locational choices of firms. It is also a period in which the Indian economy grew rapidly - but unevenly - so that regional inequalities increased, with some regions forging ahead while others have lagged behind.

There is a body of research which tries to understand this process (e.g. Desmet et al (2015), Ghani et al. (2016)). Their findings will be reviewed below, but it is fair to say that two strands have emerged: studies attempting to quantitatively measure the patterns of spatial inequalities and their changes over

[1] University of Kent, School of Economics.
[2] Asian Development Bank.
[3] Asian Development Bank.
[4] University of Kent, School of Economics.

time, and studies which try to understand their determinants. Both strands of research ask policy-relevant questions and offer some conclusions.

This study belongs to the first category and chiefly aims to quantify the extent of geographic concentration among Indian industries over the period 1998 - 2013. To do so, the study takes advantage of the recent advances in geographical indices by measuring industrial concentration with a spatially adjusted index of industrial concentration based on Ellison and Glaeser (1997) (henceforth EG). The paper's main contributions flow from the data that we marshal for the task. The data that we use in our analysis come from the Economic Censuses of India and have not yet been used to study questions of industrial agglomeration - although these data are uniquely suited for it. While most previous studies have focused their analysis on the formal manufacturing sector, our data allow us to characterize industrial concentration in the entire economy (excluding agriculture), including manufacturing *and* services, formal *and* informal sectors - at a suitably geographically disaggregated level of analysis.

This is of particular importance for a deeper understanding of the Indian economy as it is the service sector which has been driving India's fast growth rates. This is despite the fact that most of the reforms of the early 1990s were focused on unleashing India's manufacturing sector from restrictive trade and FDI policies, and an industrial licensing regime that influenced what, how much, and where firms could invest. A closer look at the spatial development of economic activity in India becomes particularly important when we consider what has not been reformed or addressed. Chief among these are labor and land regulations on the policy front, and India's large deficits in infrastructure (Panagariya 2008). Given that the impact of these constraints is concentrated in urban locations, an examination of the spatial development of economic activity can be illuminating.

Our main tool for examining the spatial development of economic activity in India is the Ellison and Glaeser (1997) index (henceforth EG index or EGI), which quantifies the degree of spatial concentration among plants in an industry. The EG index overcomes major shortcomings of the previous indices of industrial concentration such as that proposed by Krugman (1991), and remains an essential tool among economic geographers and others who study spatial patterns of economic

activity. The main limitation of most previous indices is that they fail to distinguish between spatial concentration due to industrial characteristics and spatial concentration due to agglomeration economies. The Ellison and Glaeser (EG) index controls for the industrial structure of industries and, by doing so, avoids an incorrect classification of industries as spatially concentrated when they are, for example, single-plant monopolists. A refinement to this index was developed by Guimarães et al (2011) which accounts for a well-known limitation of the EG index: it ignores the geographical positions of regions in space: the so-called 'checkerboard problem'. We use this refinement to thoroughly characterize the evolution of spatial inequalities in India at the turn of the twenty-first century and the first decade of the new millennia.

The paper's main findings can be summarized as follows. First, we show that levels of agglomeration have been higher in manufacturing industries than in service industries - and highest in those manufacturing industries that are especially capital intensive or associated with modern technology - especially information and communications technology (ICT). This finding for India is in contrast with the United States, for example, wherein the most agglomerated industries tend to be those associated with more low technology industries - in particular those associated with the first wave of the industrial revolution (e.g. textiles). This feature of India's economic landscape may be related to India's history of promoting capital and skill intensive industries (Kochhar et al, 2006).

Next, we turn to an examination of coagglomeration, in which we document similar patterns. Specifically, we show that levels of coagglomeration between pairs of industries are highest between certain high technology and skill intensive industries (e.g. manufacturing of office and computer machinery and computer related services). We interpret this finding as being indicative of substantial technology spillovers across plants in different high-tech industries.

Finally, we document that average levels of industrial concentration have been falling dramatically over time, and that this trend is driven by decreases in concentration among capital intensive sectors. This trend is in stark contrast with that of China over a similar period (Lu and Tao, 2009), where agglomeration levels seem to have increased over time. The fact that agglomeration has been decreasing in India may reflect the presence of substantial congestion costs that dissuade firms from

taking advantage of what might otherwise be efficiency enhancing levels of industrial concentration. We provide some suggestive evidence that the observed de-agglomeration may be related to congestion costs caused by inefficient land management policies.

The rest of the paper is organized as follows. First we summarize the related literature. Following this we explain the methodology we use to measure agglomeration and coagglomeration, and describe our data in detail. Then we go over the results of our analysis, and conclude with a discussion of the policy implications of our results.

2 Related literature

Following the seminal paper by Ellison and Glaeser (1997), a number of studies applied similar methodologies to examining the industrial agglomeration in different countries. For instance, Maurel and Sedillot (1999) examines the geographic concentration of French manufacturing industries in 1993 and finds similarities between France and the US regarding the most and least localized industries. The authors also find that some high technology industries are highly geographically concentrated, which they attribute to the importance of knowledge spillovers for these industries.

Studying the case of the UK in 1992, Devereux et al. (2004) find that the most agglomerated industries are those older and relatively low-tech ones and that high-tech industries are actually less agglomerated. Furthermore, they find that higher survival rates and lower entry rates are associated with the more agglomerated industries. Meanwhile, new entry could also reinforce the agglomeration of some of the most concentrated industries.

For developing countries, Lu and Tao (2009) represents the first comprehensive examination of China's industrial agglomeration. The evidence shows that the extent of geographic concentration in China's manufacturing industries, as measured by the Ellison-Glaeser index, has increased steadily between 1998 and 2005. The trend manifests encouraging effects of the economic reforms that China launched in 1978 to release market forces in resource allocation. However, comparing the degree of agglomeration in 2005 with that of developed countries in the late 1980s and 1990s, the paper finds that China's manufacturing industries still had a significantly lower level of agglomeration, which the authors argue could be explained by local protectionism across the country.

Fernandes and Sharma (2012) study the Indian case using data from the Annual Survey of Industries for the period of 1980-1999. They find that the average EG index was stable in the 1980s and decreased considerably in the 1990s. Furthermore, the paper provides evidence that de-licensing and FDI liberalization led to reduced spatial concentration of manufacturing whereas trade reforms had no significant effect.

All of the above literature focuses on industries in the formal[5] manufacturing sector. Moreover, the data used often contain a sample of relatively large firms or a census of very large firms and a sample of smaller firms. This is especially true of the papers on agglomeration in India, as they almost exclusively rely on data from the Annual Survey of Industries (ASI), which is a made up of a census of formal firms with more than 100 (or 200 - in some years) employees and a 20% sample of smaller - but still formal - firms. Unfortunately, the sample included in the ASI is only representative at the State level, and hence is not appropriate for analysis at sub-State geographic levels.

Our study also relates to another strand of the literature, which considers the relationship between industrial vintage and geographical agglomeration. Desmet and Rossi-Hansberg (2009) provide a theory of the spatial evolution of industries in which economic activity disperses as a result of technological diffusion besides the traditional agglomerative forces (i.e. input-output linkages, knowledge spillovers, and labor market pooling) and congestion costs, which lead to geographic concentration and dispersion, respectively. According to the theory, it is likely that an "old" industry will appear more geographically dispersed than a "young" industry as the technology impacting the former has been sufficiently diffused with time.

Applying the theoretical framework, Desmet et al. (2015) show that manufacturing has grown faster in districts with lower initial manufacturing density between 2000 and 2005. In contrast, districts with lower and higher service density have gained more services than the medium service density districts. The evidence - though based on an approach distinct from Ellison and Glaeser (1997) - is supportive of the theory. A related paper is Ghani et al (2016), which uses sample data for the formal manufacturing sector, the informal manufacturing sector, and some parts of the informal service sector to argue that, over the 2000s, informal service industries became more urbanized while formal

[5] By "formal" we mean those enterprises that are registered with some branch of the government.

manufacturing became less urbanized. Our own study is complementary with these others as we study a different phenomenon (i.e. industrial agglomeration rather than urbanization) using a very different set of data.

3 Methodology

Our main results make use of a spatially-adjusted version of the EG index in order to quantify the degree of geographic concentration among plants in an industry. As was briefly discussed in the introduction, the EG index takes into account the way an industry is organized by incorporating a measure of industrial structure into the index. In this section we will briefly explain the essential components of the index.[6]

The index for industry *i* in a country with *M* regions (indexed by *m*) can be expressed using vectors as

$$\gamma_i^{EG} = \frac{G_i - H_i(1 - X'X)}{(1 - H_i)(1 - X'X)} \qquad (1)$$

where H_i is a Herfindahl index measuring the industry concentration at plant level, G_i is an index of geographical concentration defined as $G_i=(S-X)'(S-X)$ where the vector S gives the fraction of employment in industry i across geographical areas m and $X'=[x_1, x_2, \ldots, x_M]$ is the vector of the aggregate employment across geographical areas m. The Herfindahl index is defined as $H_i = \sum_{j=1}^{N_i} z_j^2$ where N_i is the number of plants in industry i and z_j is the share of employment of plant j in industry i.

A limitation of the index is that it does not take into account the geographical position of regions – not even adjacent regions – even though the construction of the index requires spatial data. That is, plants are considered to be agglomerated if they are located in the same region. If they are not located in the same region, the index does not distinguish whether the plants are located in adjacent regions or in regions located on opposite sides of the country. This problem, known as the checkerboard problem,

[6] For a full account of the theoretical motivation and detailed derivation of the index, please see Ellison and Glaeser (1997).

has been addressed by Guimarães et al (2011) who developed a spatially weighted version of the index by introducing a 'neighbourhood effect' and which adjusts the EG index as follows

$$\gamma_i^{SWEG} = \frac{G_i^S - H_i(1 - X'\Psi X)}{(1 - H_i)(1 - X'\Psi X)} \qquad (2)$$

where H_i and X' are defined as in equation (1), $G_i^S=(S-X)'\Psi(S-X)$ is the spatially weighted version of the geographical concentration index and Ψ is a spatial weight matrix. Ψ is defined as $\Psi=W+I$ where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{W}$ is a matrix in which, following Guimarães et al (2011), elements representing adjacent regions are given a weight of 1 while elements representing non-adjacent regions are given a weight of 0. Note that if $\Psi=\boldsymbol{I}$ (i.e. adjacent regions are also given 0 weight), the index reduces to the standard Ellison and Glaeser measure.

Our measure of pairwise coagglomeration between industry *i* and industry *j* is given by

$$\gamma_{ij}^c = \frac{G_{ij}}{(1 - X'X)} \qquad (3)$$

where X is defined as before and G_{ij} , which captures the extent to which industries *i* and *j* co-locate, is defined as $G_{ij} = (S_i - X)'(S_j - X)$ where S and X are defined as before.

4 DATA

The main source of data we use is the Economic Census of India (EC), which is conducted by the Central Statistics Office of the Ministry of Statistics and Programme Implementation every 7-8 years. The EC is a countrywide census of establishments engaged in all economic activities except crop production and plantations.[7] In this study, we use public-use micro records from the fourth, fifth and sixth editions of the EC carried out in 1998, 2005 and 2013, respectively (henceforth EC 1998, EC 2005 and EC 2013). The data allows for the geographic location of establishments to be identified at

[7] We use the terms establishment, enterprise, and firms interchangeably.

the district and town/village level.[8] All three rounds of the economic census provide information on an establishment's number of employees and major economic activity.

The main advantages of using the EC as a data source instead of the most common alternatives spring from the fact that it is a census with almost universal scope. First, there are no concerns regarding representativeness at any geographic level. This is a significant concern with other datasets based on representative samples. For example, the ASI - the most commonly used dataset for studying spatial patterns of economic activity in India - is only representative at the state level. States in India are enormous, with 10 having populations of 50 million or above, and 3 having populations of more than 100 million. To make meaningful claims about geographic concentration, it is therefore essential to be able to conduct analysis at finer geographic levels. Relatedly, other datasets - including the ASI - suffer from changes in sampling methodology over time that make it difficult to study intertemporal trends with confidence. Second, the EC is the only dataset that includes both formal and informal activities in one dataset. Studies that focus exclusively on only the formal or the informal sectors (as nearly all previous studies have done) will therefore miss any tendency of informal establishments to agglomerate with formal establishments (or vice-versa). Finally, the EC is the only dataset that includes both manufacturing and service sectors.

In order to make the EGIs and spatially weighted EGIs comparable across years, we had to match the administrative boundaries of districts and industry classifications across the three ECs. More specifically, EC 1998, 2005 and 2013 are based on the household listings and district definitions of Population Census 1991, 2001 and 2011, and adopt National Industrial Classification of 1987, 2004 and 2008, respectively. We spent significant manual effort to account for changes in district boundaries as well as changes in industrial definitions at 3 or 4-digit levels. This effort resulted in all three EC waves assigned with districts corresponding to the 2001 Census boundaries and industry codes corresponding to the 2004 2-digit NIC codes. There are a total of 585 districts and 56 industries in the secondary and tertiary sectors on which our analysis focuses (31 industries belong to the secondary sector while 25 belong to the tertiary sector).

[8] A village is the rural counterpart to a town.

Table 1 presents counts of establishments and employment by EC and sector in India from 1998 to 2013. The number of non-agricultural establishments has increased from 26.9 million in 1998 to 35.7 million in 2005 and 45.4 million in 2013, or by 4.2% annually between 1998 and 2005 and 3.0% between 2005 and 2013. The number of tertiary firms accounts for nearly 3 quarters of the total while the growth rates are similar between the secondary and tertiary industries.

As far as employment is concerned, the total employment in the non-agricultural industries has increased on average at a 2.3% annual rate between 1998 and 2013 reaching 108.4 million in 2013. Manufacturing accounted for about one third of employment, which was higher than its share in firm count. The growth of employment between 1998 and 2005 was dominated by the tertiary sector while manufacturing showed a catch-up between 2005 and 2013.

Given that EC microdata have not previously been used to examine the spatial structure of economic activity, we undertook some comparisons with respect to sectoral totals and composition between the EC and the NSSO's EUS (National Sample Survey Organisation's Employment and Unemployment Survey). The results suggest that the sectoral distributions are reasonably close between the two datasets with the manufacturing to services employment ratios closely mirroring each other. We found that the main difference between the EC and NSSO's EUS is in the share of total employment accounted for by formal firms (which we take here to be firms with more than 10 workers, an assumption consistent with many Indian regulations, including those governing the manufacturing sector). This is likely to be due to under-coverage of the very smallest firms in the economic census (mostly own account enterprises).[9] With this caveat, the economic census appears to be a good source for studying industrial agglomeration of India over a relatively long time period. To make sure that none of our primary findings are biased by potential undercounting of the smallest enterprises, we redo some of our main analysis using only those establishments with 5 or more workers. The summary statistics for such establishments are also presented in Table 1.

[9] This is also the conclusion of Manna (2010) and Unni and Raveendran (2006).

5 Results

Our results are divided into two sections: 4.1) Industry level indices of concentration, and 4.2) Indices of coagglomeration between pairs of industries.

5.1 Industry Level Agglomeration Indices

Table 2 presents summary statistics for both weighted and unweighted EG indices over time. We see that a clear and robust pattern emerges – there has been a gradual decrease of spatial concentration for all industries from the turn of the century throughout the first decade of the twentieth century. This trend is visible in both versions of the EG index, in the simple average as well as in median values. Taking the median of the spatially weighted EGIs, we see that the decline is profound, from a high of 0.114 in 1998 to a low of 0.028 in 2013. Figure 1 plots the share of industries that may be considered "strongly clustered", where, following Ellison and Glaeser (1997), we define industries to be strongly clustered if they have an EG index value of .05 or above. However one chooses to look at the data, the trend is clear: levels of concentration have been falling dramatically over a period that coincides with particularly fast growth in India.[10]

However, these economy-wide averages can mask sectoral differences. Therefore Figures 2 and 3 plot the median values of spatially weighted and unweighted indices separately by major sector, distinguishing industries in the secondary sector from those in the tertiary sector. Several facts are apparent from these figures. First, levels of geographic concentration have been systematically higher in the secondary sector than in the tertiary sector. Second, while both sectors have experienced a

[10] As we noted earlier, one potential weakness of the Economic Census is that it may undercount the very smallest informal establishments. To make sure that changes in the degree of undercounting over time are not driving any of our intertemporal results, we include the following robustness check: we re-generate the EG Indices omitting all establishments with less than 5 workers, since it is establishments of this size that are potentially undercounted. Doing so will eliminate this potential source of measurement error. The results, displayed in Table 12, Figure 4 and Figure 5 show that the intertemporal patterns observed in our main results are not sensitive to the inclusion or exclusion of the very smallest establishments.

general decline in average levels of concentration, the biggest portion of the economy-wide decline in concentration is accounted for by the decline in concentration within the secondary sector.[11]

It is also illuminating to break up the results by capital intensity of industries within the secondary sector. The results, depicted in Figures 6 and 7, show that the drop in concentration seems to have been driven by a precipitous drop in concentration among capital intensive firms in particular. This result is most stark when looking at the spatially weighted version of the index (Figure 7).

We can obtain further insights into the patterns of the spatial distribution of industries and possible suggestions of their causes by looking at the top and the bottom of the distribution of the spatially-weighted EG (SWEG) index. We present the results in six tables. Each table shows either the top ten or the bottom ten industries in the years 1998 and 2013 (i.e. the endpoints of our data) for the entire economy, manufacturing industries, and services respectively.

Examining the table with the list of the top most concentrated industries, we see that manufacturing industries dominate the list, a fact consistent with Figures 2 and 3. Indeed, there are only two service-sector industries, and neither of them is persistently present in the top-ten list between 1998 and 2013. The top-ten manufacturing industries show much more persistence meaning that there are quite a few of them which were among the most concentrated in 1998 and remained so in 2013 as well. For example, the industries of electrical machinery, rubber and plastic, motor vehicles, and radio, television and communication respectively are persistently among the most concentrated industries. Overall, almost all of the industries in that group are high value-added industries with the technology of the 'second industrial revolution' and ICT era. Looking at the top most localized manufacturing industries only confirms that picture, which is not surprising since they dominate the list of the most localized industries altogether. Services show an interesting pattern. The top ten most localized services are ICT related industries such as computer related industries, and post and telecommunication. Airport transport as well as water transport belong to that group as well with the

[11] Although the weighted and unweighted versions of the indices agree on the general trend, they differ somewhat in the trajectory and and magnitude of the decline. In particular, the secondary sector experiences a decline and a plateau since the early 2000s in the unweighted version of the index while the weighted version shows an accelerating decline since the early 2000s.

research and development sector joining them in 2013. Otherwise, most of the most localized services in 1998 were so in 2013 as well. Overall, we can detect a persistency of industries at the top of the distribution.

Looking at the least spatially concentrated industries reveals that they are dominated by services, a picture again consistent with Figure 3. Indeed, there is only one manufacturing industry in that category – manufacture of food products and beverages – and only in the year of 1998. The least localized manufacturing industries are clearly those which can be characterized as traditional industries – or industries of the first industrial revolution – such as textile, apparel, food products, metal products, though we also find an industry of the second industrial revolution – chemical products, or petroleum refinement. The least localized services are those related to retailing (e.g. sales of motor vehicles), and (still) labor intensive and high transport costs services such as hotels and restaurants, education, and repairs respectively. These are also persistently the least localized services between 1998 and 2013.

Overall, we can say that the top and bottom distribution of industries reveal a persistent presence of the same industries. The top ten of the most localized manufacturing industries are mostly related to the electrical industries, transportation industries, and the industries of the ICT era, while the least localized industries are mostly what we called traditional industries of textile, apparel, and metals. Services exhibit a similar pattern to a certain extent. The top localized services are related to the ICT era while the least localized are those which can be characterized as traditional.

5.2 Coagglomeration

We can gain further insights into the spatial distribution of manufacturing and services by looking at pairwise coagglomeration indices. Table 11 presents the top most coagglomerated industries and a very clear pattern emerges: a strong coagglomeration of high-tech, ICT-based manufacturing *and* service industries. That pattern is even more prevalent in 2013 than in 1998. In other words, it is not only that establishments in the same high technology manufacturing industry agglomerate together. In

addition, establishments in high technology industries co-agglomerate with establishments from different - but still high tech industries. The strength of this pattern is highly suggestive of reasonably broad knowledge or technology spillovers across even 2 digit industries.

6 Discussion

In this paper we use uniquely suited data to generate a set of basic facts, patterns and trends regarding the degree of spatial concentration of economic activity in India, in both manufacturing and service sectors. The most striking findings are the following. First, levels of agglomeration in India seem to be highest in capital intensive and relatively modern / ICT-heavy industries. This is stark contrast with the US, for example, in which the most agglomerated industries are typically those from the first wave of the industrial revolution such as textiles (Ellison and Glaeser, 1997). This finding may be particularly interesting given India's history of emphasizing capital intensive industries in its development policy (Kochhar et al, 2006).

Second, we find that plants in modern / high-tech industries tend not only to agglomerate together, they also co-agglomerate with plants in *other* modern / high-tech industries. This fact may be indicative of the existence and reach of knowledge and technology spillovers across 2-digit industries.

Finally, perhaps our most interesting finding is that industrial agglomeration has decreased substantially between 1998 and 2013 - with the bulk of the drop taking place between 1998 and 2005. The decline seems to be part of a trend that started in the early 1990s, when reforms were put in place that involved dismantling India's policy of industrial licensing (known as industrial de-licensing) along with liberalization of FDI policies—which Fernandes and Sharma (2012) find to be associated with declining agglomeration levels. This stands in sharp contrast with the movement toward higher industrial spatial concentration in China over a comparable time period. Interestingly, the recent trend we document in India has been driven in large part by a decrease in the spatial concentration of capital intensive industries - i.e. those industries that were initially the most spatially concentrated -, which may reflect a change in India's historical policy of prioritizing capital intensive industries.

The trend is important for three reasons. First, there is a good deal of research showing that agglomeration of economic activity plays an important role in boosting firm productivity and economic growth (e.g. Duranton 2014). Second, with the caveat that the EG indices may not be reliable when compared across countries, the literature documents substantially higher EG indices in developed countries than those we have estimated for India. Third, China, which has been growing faster since the 1980s, has experienced increased industrial agglomeration. The lack of spatial agglomeration in India suggests that an important source of economic growth seems to be missing.

Together, these facts imply that the forces of congestion are preventing a more efficient level of agglomeration from taking place. Agglomeration results from firms' desires to collocate with others and thereby exploit input-output linkages, thicker local labor markets, and knowledge spillovers. They are the reason for a close link between urbanization and industrialization. In India, however, several factors seem to be contributing to high congestion costs in cities, i.e., the locations one would expect to experience industrial agglomeration. These include underinvestment in urban infrastructure (according to McKinsey Global Institute (2010), Indian cities have been spending only around $17 per capita on urban infrastructure versus benchmark needs estimated at $100 per capita) and restrictive regulations on land use and building regulations, including very low floor area ratios, that have made urban land artificially scarce (Sridhar, 2010 and Brueckner and Sridhar, 2012). These factors are likely to have raised the costs of transport and (renting or buying) land and thus encouraged the spread of industry.

As a simple and informal test of how a bad policy environment may be driving the observed trends, we compute and compare the EG Index for two separate parts of India: one part composed of states which have had relatively good land management policies; and a second part composed of states with relatively bad land management policies. Our partition of states into the two categories is based on Hasan, Jiang, and Kundu (forthcoming).[12] Strictly speaking, the fact that Indian firms may be able to

[12] The partition divides states according their success on 7 different measures that are relevant to urban land management. These 7 measures include 1) undertaking reforms in rent control; 2) repealing the urban land ceiling reform act (ULCRA); 3) earmarking 25% of developed land in all housing projects for low income groups; 4) achieving 85% coverage and 90% collection efficiency of property taxes; 5) reducing stamp duty to 5 percent

locate across the two sets of states may be potentially problematic when computing the EG index separately for each part of India as though they were different countries. However, to the extent that any bias induced by firms' moving across state lines is not large, the results - displayed in Table 13 - are suggestive. We find that states with bad land management started off in 1998 with much higher concentration than in states with flexible regulations. Over time, however, concentration in those states fell - so much so that average concentration is now lower in bad management states than in good ones. Overall, these findings are consistent with the hypothesis that poor land management policies contribute to congestion costs and may thus drive the significant trend towards de-agglomeration.

Another reason for the relatively low degree of industrial agglomeration may be related to the size distribution of Indian firms. Lafourcade and Mion (2007) show that large plants are more spatially concentrated than small ones. The scale of firms is relatively small in India (Hasan and Jandoc 2010). Moreover, as shown in Table 1, growth in the number and employment of firms with 5 or more employees lagged behind overall growth from 1998 to 2005. The majority of the increase in firms and employment over this period came from micro enterprises, which may have been more likely to geographically dispersed than larger establishments. Policy distortions such as restrictive labor regulations that prevent firms from growing larger could also discourage firms from moving closer together, because the enforcement of labor and other regulations is quite likely to be weaker the farther away one is from established urban centers, providing firms another reason to spread their plants.

This paper has aimed to generate some basic facts and patterns regarding the spatial concentration of economic activity in India. The findings are suggestive of potential policy implications, but they are far from definitive. In future work we hope to shed light on the particular determinants of agglomeration and co-agglomeration of manufacturing as well as service industries in India. In

---

or less; 6) simplifying the legal and procedural framework for converting agricultural land for non-agricultural purposes; and 7) introducing a computerized process of registration of land and property.

particular, we hope to test our hypotheses that congestion costs in India are disrupting the forces of agglomeration, and that these costs may be related to specific economic policies.

References

Ellison, G., Glaeser, E.L. and Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), pp.1195-1213.

Brueckner, J. K., and Sridhar, K. S. 2012. Measuring welfare gains from relaxation of land-use restrictions: The case of India's building-height limits. *Regional Science and Urban Economics,* 42(6): 1061-1067.

Desmet, K. and Rossi-Hansberg, E., 2009. Spatial growth and industry age. *Journal of Economic Theory*, *144*(6), pp.2477-2502.

Desmet, K., Ghani, E., O'Connell, S. and Rossi-Hansberg, E., 2015. The spatial development of India. *Journal of Regional Science*, *55*(1), pp.10-30.

Devereux, M.P., Griffith, R. and Simpson, H., 2004. The geographic distribution of production activity in the UK. *Regional science and urban economics*, *34*(5), pp.533-564.

Duranton, Gilles. 2015. "Growing Through Cities in Developing Countries." *The World Bank Research Observer* 30 (1): 39–73.

Fernandes, A.M. and Sharma, G., 2012. Together we stand? agglomeration in Indian manufacturing. World Bank Working Paper No. 6062.

Ghani, E., Goswami, A. G., and Kerr, W., 2016. Spatial Development and Agglomeration Economies in Services: Lessons from India. World Bank Policy Research Working Paper 7741.

Hasan, Rana and Karl Robert L. Jandoc. 2010. “The Distribution of Firm Size in India: What Can Survey Data Tell Us?” ADB Economics Working Paper #213.

Hasan, R., Jiang, Y. and Kundu, D. (Forthcoming). The Growth of Indian Cities and “Good” Jobs: Evidence from the 2000s. *India Policy Forum.* NCAER. Sage Publications.

Kochhar, K., Kumar, U., Rajan, R., and Subramanian, A., and Tokatlidis, I., 2006. "India's pattern of development: What happened, what follows?," *Journal of Monetary Economics*, 53(5), pp. 981-1019.

Lafourcade, M. and Mion, G., 2007. Concentration, agglomeration and the size of plants. *Regional Science and Urban Economics*, *37*(1), pp.46-68.

Lu, J. and Tao, Z., 2009. Trends and determinants of China's industrial agglomeration. *Journal of urban economics*, *65*(2), pp.167-180.

Manna, G. C., 2010. Current Status of Industrial Statistics in India: Strengths and Weaknesses. *Economic and Political Weekly*, XLV(46), pp. 67-76.

Maurel, F. and Sédillot, B., 1999. A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, *29*(5), pp.575-604.

McKinsey Global Institute. 2010. “India’s urban awakening: Building inclusive cities sustaining economic growth.” New York: McKinsey & Company.

Sridhar, K. S. (2010a). “Impact of land use regulations: Evidence from India’s cities.” *Urban Studies*, 47(7): 1541-1569.

Unni, J., and Raveendran, G., 2006. Manna, G. C., 2010. Are the Results of the Economic Census Robust? *Economic and Political Weekly*, XLI(33), pp. 3558-3560.

Figures and Tables

[Attached separately]

Figures and Tables

Figure 1: Share of Industries Strongly Clustered (EGI $>$ .05) Over Time

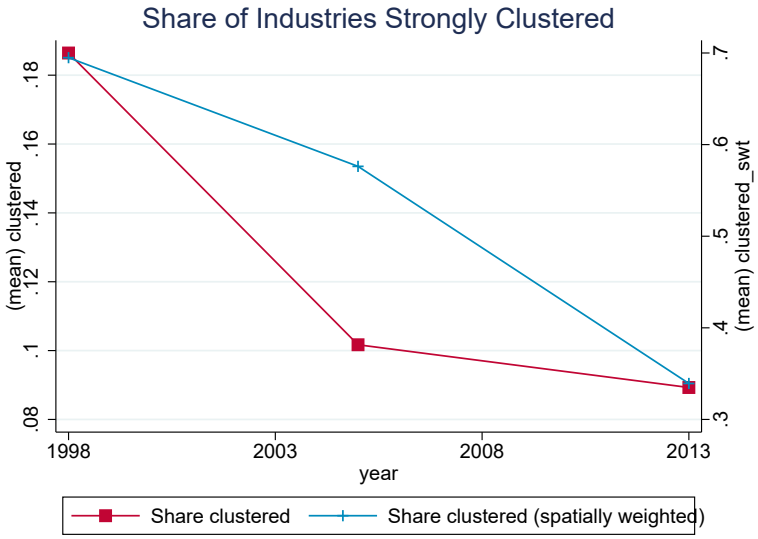

Figure 2: Median EGIs Over Time by Sector

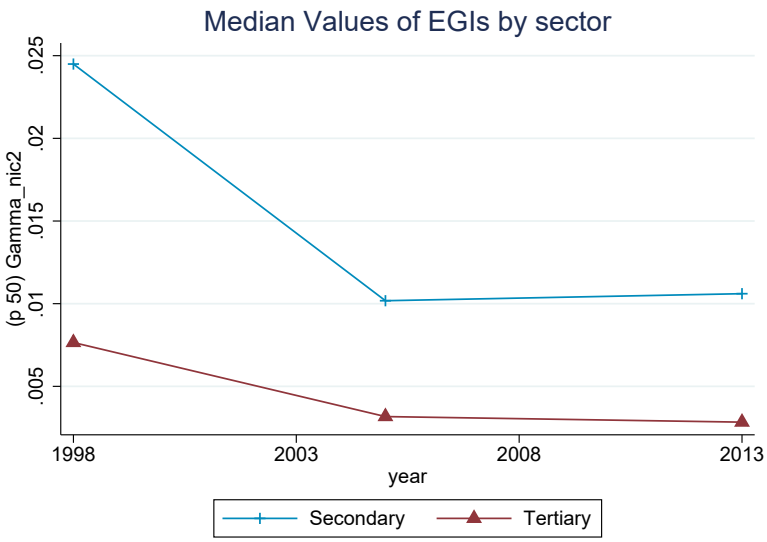

Figure 3: Median Spatially Weighted EGIs Over Time by Sector

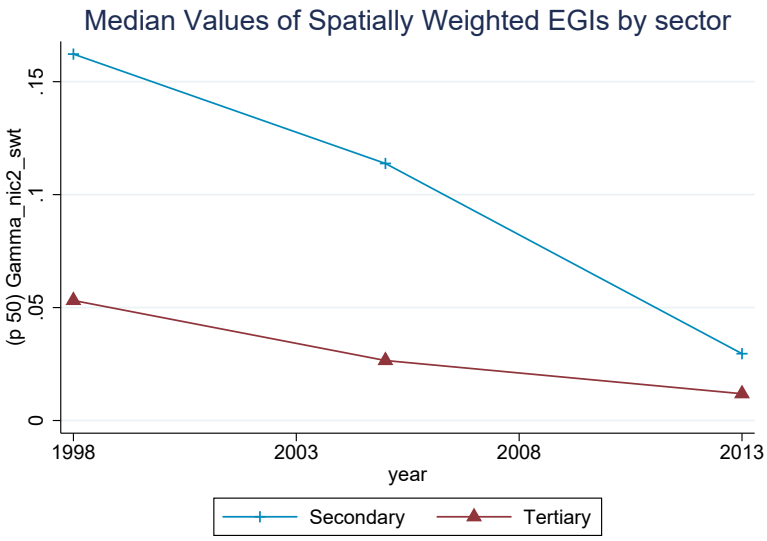

Figure 4: Share of Industries Strongly Clustered (EGI $>$ .05) Over Time (excluding establishments with $<$ 5 workers)

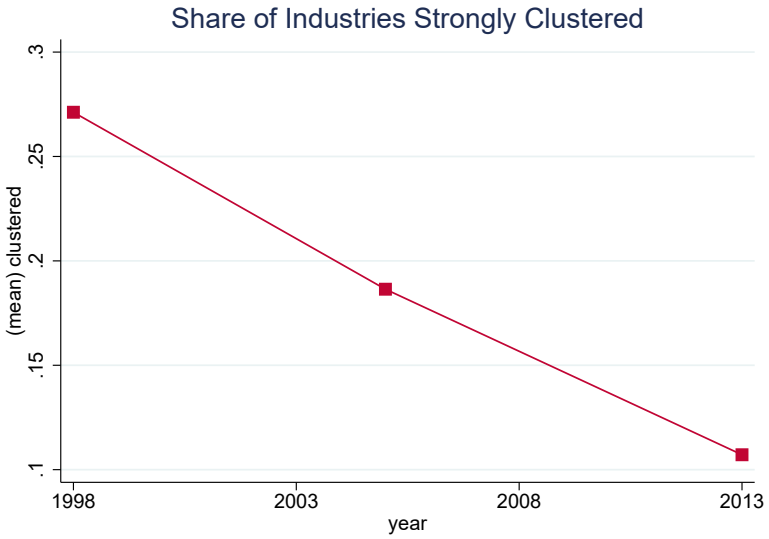

Figure 5: Median EGIs Over Time by Sector (excluding establishments with $<$ 5 workers)

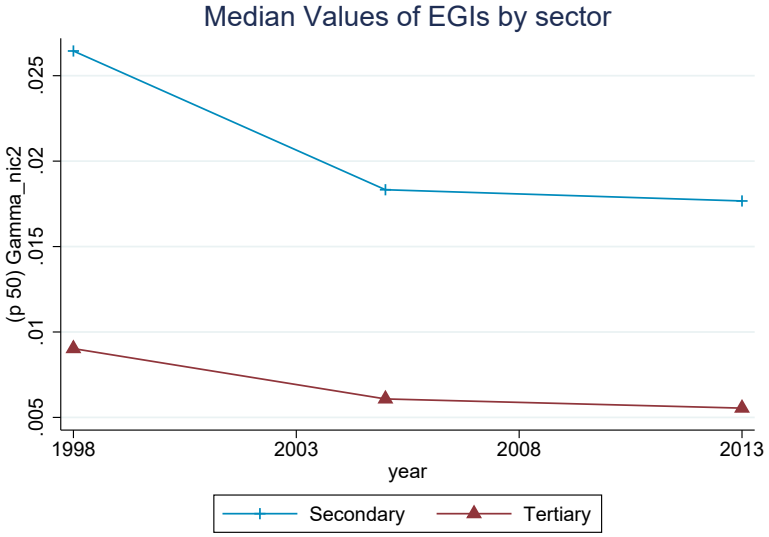

Figure 6: Median EGIs Over Time by Capital Intensity of Industry

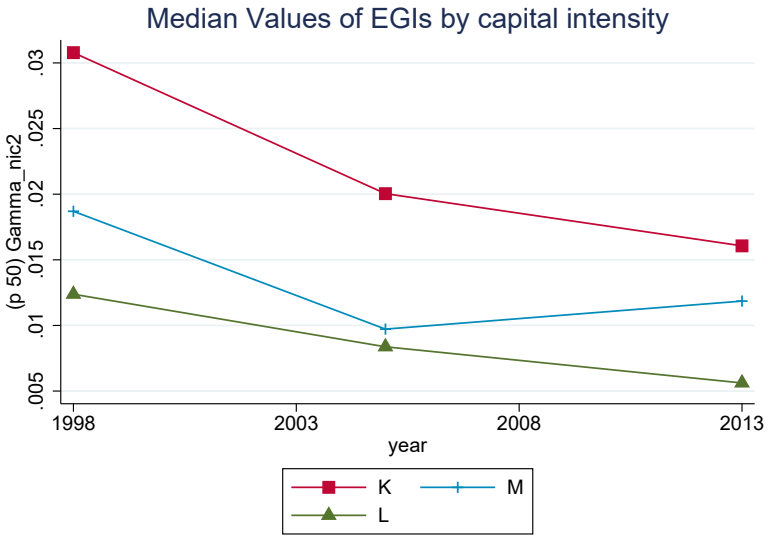

Figure 7: Median Spatially Weighted EGIs Over Time by Capital Intensity of Industry

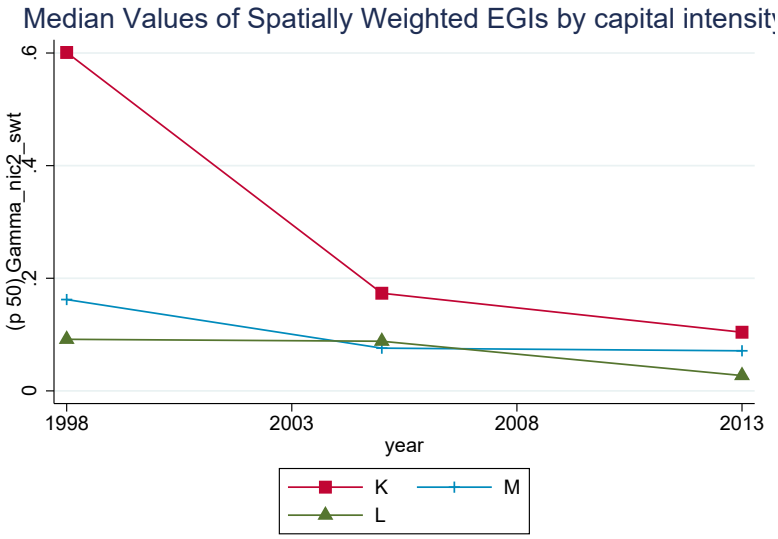

Table 1: Economic Census (summary of enterprise and employment data)

	All Firms			Firms with >=5 employees		
	1998	**2005**	**2013**	**1998**	**2005**	**2013**
# Firms						
Secondary	6,859,307	8,792,537	11,604,026	886,688	906,852	1,173,329
% annual growth		3.6	3.5		0.3	3.3
% annual growth between 1998 and 2013			3.6			1.9
Tertiary	20,003,166	26,954,469	33,759,760	1,491,485	1,795,107	2,348,099
% annual growth		4.4	2.9		2.7	3.4
% annual growth between 1998 and 2013			3.6			3.1
Total	26,862,473	35,747,006	45,363,786	2,378,173	2,701,959	3,521,433
% annual growth		4.2	3.0		1.8	3.4
% annual growth between 1998 and 2013			3.6			2.7
# Employment						
Secondary	25,883,567	27,245,289	34,223,108	15,156,137	13,994,749	17,413,457
% annual growth		0.7	2.9		-1.1	2.8
% annual growth between 1998 and 2013			1.9			0.9
Tertiary	50,637,754	62,745,231	74,188,259	21,667,201	24,474,130	25,988,389
% annual growth		3.1	2.1		1.8	0.8
% annual growth between 1998 and 2013			2.6			1.2
Total	76,521,321	89,990,520	108,411,367	36,823,338	38,468,878	43,401,850
% annual growth		2.3	2.4		0.6	1.5
% annual growth between 1998 and 2013			2.3			1.1

Table 2: Summary Statistics for EGIs Over Time

	count	mean	sd	min	p50	max
1998						
Gamma (unweighted)	59	0.035	0.063	0.000	0.015	0.367
share >.05 (unweighted)	59	0.186	0.393	0.000	0.000	1.000
Gamma (spatially weighted)	59	0.386	0.783	0.003	0.114	4.876
share >.05 (spatially weighted)	59	0.695	0.464	0.000	1.000	1.000

	count	mean	sd	min	p50	max
2005						
Gamma (unweighted)	59	0.017	0.018	0.000	0.009	0.084
share >.05 (unweighted)	59	0.102	0.305	0.000	0.000	1.000
Gamma (spatially weighted)	59	0.111	0.131	0.001	0.061	0.626
share >.05 (spatially weighted)	59	0.576	0.498	0.000	1.000	1.000

	count	mean	sd	min	p50	max
2013						
Gamma (unweighted)	56	0.018	0.029	0.000	0.007	0.185
share >.05 (unweighted)	56	0.089	0.288	0.000	0.000	1.000
Gamma (spatially weighted)	56	0.088	0.170	0.001	0.028	0.895
share >.05 (spatially weighted)	56	0.339	0.478	0.000	0.000	1.000

Table 3: Values of EGI for several significant industrial categories over time:

NIC Code	Industry	spatially weighted	1998	2005	2013
15	Manufacture of food products and beverages	no	.004	.002	.002
		yes	.013	.013	.012
17	Manufacture of textiles	no	.008	.011	.011
		yes	.021	.32	.024
27	Manufacture of basic metals	no	.013	.008	.007
		yes	.127	.061	.017
34	Manufacture of motor vehicles ...	no	.033	.030	.052
		yes	**.978**	.162	.301
40	Electricity, gas, steam and hot water supply	no	.049	.002	.001
		yes	**2.062**	.015	.005
45	Construction	no	.003	.003	.008
		yes	.012	.012	.011
52	Retail trade, except of motor vehicles ...	no	.000	.000	.000
		yes	.006	.001	.001
72	Computer and related activities	no	.051	.038	.034
		yes	**.653**	**.151**	**.052**

Table 4: Highest Gammas (most localized)

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
30	Manufacture of office, accounting and cpu machinery	4.876	0.138
31	Manufacture of electrical machinery ...	2.428	0.067
40	Electricity, gas, steam and hot water supply	2.062	0.049
32	Manufacture of radio, tv and comm. equipment	1.768	0.058
41	Collection, purification and distribution of water	1.322	0.037
22	Publishing, printing and reproduction of recorded media	1.255	0.036
34	Manufacture of motor vehicles ...	0.978	0.033
25	Manufacture of rubber and plastics products	0.880	0.029
72	Computer and related activities	0.653	0.051
18	Manufacture of wearing apparel ...	0.529	0.015
2013			
32	Manufacture of radio, tv and comm. equipment	0.895	0.050
25	Manufacture of rubber and plastics products	0.779	0.026
21	Manufacture of paper and paper products	0.437	0.017
19	Tanning and dressing of leather, etc	0.356	0.031
34	Manufacture of motor vehicles ...	0.301	0.052
13	Mining of metal ores	0.236	0.185
31	Manufacture of electrical machinery ...	0.178	0.013
63	Supporting and auxiliary transport activities, etc	0.138	0.007
2	Forestry, logging, etc	0.129	0.059
16	Manufacture of tobacco products	0.124	0.062

Table 5: Highest Gammas (most localized) in Manufacturing

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
30	Manufacture of office, accounting and cpu machinery	4.876	0.138
31	Manufacture of electrical machinery ...	2.428	0.067
32	Manufacture of radio, tv and comm. equipment	1.768	0.058
22	Publishing, printing and reproduction of recorded media	1.255	0.036
34	Manufacture of motor vehicles ...	0.978	0.033
25	Manufacture of rubber and plastics products	0.880	0.029
18	Manufacture of wearing apparel ...	0.529	0.015
37	Recycling	0.434	0.031
33	Manufacture of medical, precision and optical instruments	0.322	0.020
19	Tanning and dressing of leather, etc	0.288	0.023
2013			
32	Manufacture of radio, tv and comm. equipment	0.895	0.050
25	Manufacture of rubber and plastics products	0.779	0.026
21	Manufacture of paper and paper products	0.437	0.017
19	Tanning and dressing of leather, etc	0.356	0.031
34	Manufacture of motor vehicles ...	0.301	0.052
31	Manufacture of electrical machinery ...	0.178	0.013
16	Manufacture of tobacco products	0.124	0.062
28	Manufacture of fabricated metal products, except machinery and equipment	0.124	0.007
22	Publishing, printing and reproduction of recorded media	0.092	0.006
35	Manufacture of other transport equipment	0.085	0.081

Table 6: Highest Gammas (most localized) in Services

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
72	Computer and related activities	0.653	0.051
63	Supporting and auxiliary transport activities etc	0.512	0.018
61	Water transport	0.395	0.303
99	Extraterritorial organizations and bodies	0.387	0.022
70	Real estate activities	0.234	0.010
64	Post and telecommunications	0.192	0.011
62	Air transport	0.169	0.101
67	Activities auxiliary to financial intermediation	0.129	0.022
50	Sale, maintenance and repair of motor vehicles and motorcycles	0.114	0.004
74	Other business activities	0.085	0.033
2013			
63	Supporting and auxiliary transport activities etc	0.138	0.007
70	Real estate activities	0.082	0.006
73	Research and development	0.054	0.028
72	Computer and related activities	0.052	0.034
61	Water transport	0.049	0.019
74	Other business activities	0.041	0.007
64	Post and telecommunications	0.030	0.002
62	Air transport	0.023	0.018
51	Wholesale trade, except of motor vehicles and motorcycles	0.018	0.002
90	Sewage and refuse disposal, sanitation and similar activities	0.017	0.009

Table 7: Lowest Gammas (least localized)

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
85	Health and social work	0.003	0.001
55	Hotels and restaurants	0.003	0.001
93	Other service activities	0.003	0.001
92	Recreational, cultural and sporting activities	0.004	0.001
71	Renting of machinery and equipment	0.005	0.002
52	Retail trade, except of motor vehicles and motorcycles	0.006	0.000
60	Land transport; transport via pipelines	0.008	0.004
45	Construction	0.012	0.003
66	Insurance and pension funding	0.013	0.010
15	Manufacture of food products and beverages	0.013	0.004
2013			
52	Retail trade, except of motor vehicles and motorcycles	0.001	0.000
93	Other service activities	0.001	0.001
55	Hotels and restaurants	0.001	0.001
92	Recreational, cultural and sporting activities	0.004	0.001
40	Electricity, gas, steam and hot water supply	0.005	0.001
71	Renting of machinery and equipment	0.005	0.002
85	Health and social work	0.005	0.001
60	Land transport; transport via pipelines	0.005	0.001
80	Education	0.006	0.001
50	Sale, maintenance and repair of motor vehicles and motorcycles	0.007	0.001

Table 8: Lowest Gammas (least localized) in Manufacturing

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
15	Manufacture of food products and beverages	0.013	0.004
17	Manufacture of textiles	0.021	0.008
36	Manufacture of furniture; manufacturing n.e.c.	0.028	0.008
26	Manufacture of other non-metallic mineral products	0.048	0.004
24	Manufacture of chemicals and chemical products	0.055	0.026
20	Manufacture of wood and of products of wood, except furniture	0.074	0.018
28	Manufacture of fabricated metal products, except machinery and equipment	0.109	0.010
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.112	0.086
21	Manufacture of paper and paper products	0.124	0.007
27	Manufacture of basic metals	0.127	0.013
2013			
15	Manufacture of food products and beverages	0.012	0.002
36	Manufacture of furniture; manufacturing n.e.c.	0.012	0.004
27	Manufacture of basic metals	0.017	0.007
24	Manufacture of chemicals and chemical products	0.018	0.010
33	Manufacture of medical, precision and optical instruments	0.019	0.015
26	Manufacture of other non-metallic mineral products	0.022	0.006
17	Manufacture of textiles	0.024	0.011
18	Manufacture of wearing apparel; dressing and dyeing of fur	0.026	0.003
23	Manufacture of coke, refined petroleum products and nuclear fuel	0.026	0.009
20	Manufacture of wood and of products of wood, except furniture	0.029	0.005

Table 9: Lowest Gammas (least localized) in Services

NIC 2004	Industry	Gamma (SW)	Gamma
1998			
85	Health and social work	0.003	0.001
55	Hotels and restaurants	0.003	0.001
93	Other service activities	0.003	0.001
92	Recreational, cultural and sporting activities	0.004	0.001
71	Renting of machinery and equipment	0.005	0.002
52	Retail trade, except of motor vehicles and motorcycles	0.006	0.000
60	Land transport; transport via pipelines	0.008	0.004
66	Insurance and pension funding	0.013	0.010
75	Public administration and defence	0.023	0.002
80	Education	0.026	0.001
2013			
52	Retail trade, except of motor vehicles and motorcycles	0.001	0.000
93	Other service activities	0.001	0.001
55	Hotels and restaurants	0.001	0.001
92	Recreational, cultural and sporting activities	0.004	0.001
71	Renting of machinery and equipment	0.005	0.002
85	Health and social work	0.005	0.001
60	Land transport; transport via pipelines	0.005	0.001
80	Education	0.006	0.001
50	Sale, maintenance and repair of motor vehicles and motorcycles	0.007	0.001
65	Financial intermediation, except insurance and pension funding	0.009	0.004

Table 10: Pairwise Coagglomeration between Secondary and Tertiary Sectors

Industry 1	Industry 2	Coagglomeration
1998		
Secondary	Tertiary	-.0004468
2005		
Secondary	Tertiary	-.0001629
2013		
Secondary	Tertiary	1.19e-06

Table 11: Most Highly Coagglomerated 2 digit Industries (Pairwise)

Industry 1	Industry 2	Coagglomeration
1998		
Water Transport (61)	Air Transport (62)	.2878174
Mining of uranium and thorium ores (12)	Manufacture of motor vehicles... (34)	.1148259
Manufacture of office... and cpu machinery (30)	Manufacture of electrical machinery ... (31)	.0906983
Extraction of oil and natural gas (11)	Manufacture of coke, refined petroleum products and nuclear fuel (23)	.0890916
Air Transport (62)	Other business activities (74)	.0858767
Water Transport (61)	Other business activities (74)	.0832694
Manufacture of office... and cpu machinery(30)	Manufacture of radio, tv and comm. equipment (32)	.0826168
Manufacture of office... and cpu machinery (30)	Electricity, gas and water (40)	.0798371
Publishing, printing and reproduction of recorded media (22)	Manufacture of office... and cpu machinery (30)	.0657404
Manufacture of office... and cpu machinery (30)	Manufacture of motor vehicles... (34)	.0631796
2013		
Manufacture of office... and cpu machinery (30)	Manufacture of radio, tv and comm. equipment (32)	.0629072
Manufacture of office... and cpu machinery (30)	Computer and related activities (72)	.0411532
Manufacture of motor vehicles... (34)	Manufacture of other transport equipment (35)	.03403
Manufacture of office... and cpu machinery (30)	Research and development (73)	.0271653
Manufacture of machinery and equipment n.e.c. (29)	Manufacture of office... and cpu machinery (30)	.026901
Manufacture of rubber and plastics products (25)	Manufacture of radio, tv and comm. equipment (32)	.0258752
Computer and related activities (72)	Research and development (73)	.02223
Manufacture of office... and cpu machinery (30)	Manufacture of motor vehicles... (34)	.0207206
Manufacture of radio, tv and comm. equipment (32)	Manufacture of motor vehicles... (34)	.0207108
Manufacture of other transport equipment (35)	Recycling (37)	.0195396

Table 12: Summary Statistics for EGIs Over Time (excluding establishments with < 5 workers)

	count	mean	sd	min	p50	max
1998						
Gamma (unweighted)	59	0.046	0.077	0.002	0.021	0.448
share >.05 (unweighted)	59	0.271	0.448	0.000	0.000	1.000
	count	mean	sd	min	p50	max
2005						
Gamma (unweighted)	59	0.027	0.029	0.001	0.016	0.107
share >.05 (unweighted)	59	0.186	0.393	0.000	0.000	1.000
	count	mean	sd	min	p50	max
2013						
Gamma (unweighted)	56	0.024	0.040	0.001	0.013	0.274
share >.05 (unweighted)	56	0.107	0.312	0.000	0.000	1.000

Table 13: Average EGIs in Good vs Bad Land Management Regions

	Good Land Management		**Bad Land Management**	
Year	**Mean**	**SD**	**Mean**	**SD**
1998	.07	.195	.111	.174
2005	.053	.072	.058	.073
2013	.065	.102	.051	.063