Indian Public Policy Review 2023, 4(2): 55-74
https://doi.org/10.55763/ippr.2023.04.02.003

IPPR

Nature of agglomeration of the manufacturing sector –a study of Indian Districts

Piyali Majumder[*#]

Abstract

The paper estimates the degree of agglomeration of the Indian organised manufacturing sector and examines its evolution pattern across districts over the period 2000-01 to 2009-10. The estimation of the degree of industrial agglomeration is based on plant-level data from the Annual Survey of Industries. The paper uses the spatially-weighted Ellison Glaser Index to control for the inter-district spillover effect. The overall degree of agglomeration has been moderate and, over time, it registered a declining trend. While analysing the nature of industrial agglomeration, it has been observed that most of the low-tech and medium-low-tech industries are found to be highly agglomerated. 42% of the highly agglomerated industries are also highly polluting in nature. During the period 2000-01 to 2009-10, the second-tier cities observed a rise in the number of plants belonging to the polluting industries. High-tech industries are found to be concentrated in the already industrialised states. In contrast to this, the medium-high-tech industries have been spreading across districts. The distribution of low-tech industries is found to be even across the districts.

Keywords: agglomeration economies, manufacturing, India

JEL Codes: R11; R12

Publication Date: 31 March 2023

[*] Piyali Majumder is an Associate Fellow at National Council of Applied Economic Research (NCAER).
The author is deeply grateful to Professor Aparna Sawhney, Jawaharlal Nehru University for her insights and guidance.

Introduction

While Indian manufacturing posted an impressive average annual growth rate of more than 7% during the last two decades,[1] the dilemma of achieving manufacturing industry-based economic growth -- as well as its adverse impact on the environment -- has remained an area of concern. Another major challenge in India has been to ensure balanced regional development of industries, such that growth is not restricted to a few districts, but rather has a wide geographical spread. It is critical that the poorer regions are also part of the higher productivity and income growth story, in order to achieve inclusive development.

Regional industrial concentration has persisted in India, despite overhauling of the industrial licensing regime with systematic de-regulation and liberalization since 1991. Even enhanced government capital expenditure has failed to break the trend of escalating regional inequality, although it has had a significant positive impact on the manufacturing output growth of the poorer states (Barua and Sawhney 2015). Absolute income divergence exists at both the state and district levels; however, there is evidence of higher transitional growth in districts with proximity to urban agglomeration –indicating the importance of this geographical factor in development (Das et al 2015).

Agglomeration economies have been found to be significant in the Indian manufacturing sector. The presence of intra-industry spill overs; inter-industry linkages; availability of infrastructural facilities including transport infrastructure (ensuring easy access to input and output markets), electricity, water, etc.; and government policies have acted as centripetal forces in reinforcing agglomeration of industries in Indian organised manufacturing sector (Lall et al 2004).

High-tech industries (e.g. manufacturing of machinery equipment, manufacturing of electronics and computer equipment) are concentrated mostly in urban areas, as opposed to the low-end manufacturing industries (food and beverages, leather processing, and tobacco industries). The high-tech innovative industries have a greater ability to pay high wages and land rents prevailing in densely populated urban areas compared to the low-end manufacturing industries (Lall et al 2004). The externalities arising from the availability of infrastructural facilities, large consumer markets, and the presence of a diversified industrial base or cross-industry economies, have all had a significant positive impact on the productivity of these high-tech industries (Lall et al 2004).

On the other hand, low-end manufacturing industries like food and beverages, leather processing, and tobacco industries have mostly benefitted from within-industry economies (e.g. industry-specific labour pool and technical know-how), and are located in rural areas of the country (Lall et al 2004). While analysing the intra-industry agglomeration pattern of the manufacturing sector, Mukim (2014) found that within an industry there is a close association between formal and informal sector firms. She observed that informal firms are sellers of material inputs and labour to the formal firms within an industry.

There is also evidence that organised manufacturing has begun to spread towards peri-urban and rural areas in India (Ghani et al 2012, Colmer 2014). Moreover, the development of industrial zones, industrial corridors, parks, and financial assistance in the backward regions has led to faster growth of low-density manufacturing districts, as compared to the high-density manufacturing districts in metropolitan areas (Ghani 2014).

The overall expansion of manufacturing activities in India has also raised serious concerns about the environmental problems associated with it. In 2009, the Central Pollution Control Board noted that 48% of the industrial clusters in the country were critically polluted (43 clusters out of 88 clusters monitored) as per its Comprehensive Environmental Pollution Index (CEPI). The CEPI scores, based on several pollutants, reflect the underlying environmental quality of air, surface water, and groundwater.[2] Over the years, the index scores show that pollution in industrial clusters has significantly aggravated in several areas over the years (CPCB 2009, 2011, 2013).

There have been some attempts to prepare a comprehensive environmental mapping for the location of industries ('Zoning Atlas for siting industries') across all districts by the Central and State Pollution Control Boards, on the basis of input requirements (availability of raw materials, labour, water and power supply) and environmental factors (i.e. air, the water quality of a location) in the siting of new industrial units,[3] in order to ensure that these would be economically and environmentally viable. However, results have been mixed, as compliance with the industry-specific emission standards is monitored by the State Pollution Control Board (SPCBs), and the degree of enforcement of environmental laws varies across states.

The analysis of the nature of industrial agglomeration in the existing literature has distinguished between high-tech versus low-tech industries (based on capital intensities of the production process). Less attention has been paid to the polluting nature of industries. While the traditional capital-intensive industries (manufacturing of iron and steel, chemicals, and allied products) are pollution-intensive, all capital-intensive industries are not polluting, for instance, manufacturing of electronics and electrical goods, medical and surgical instruments etc.

There are also certain labour-intensive industries, like the manufacturing of leather, plastic etc., which are highly polluting. Moreover, existing studies on agglomeration across industries have been done at an aggregated level i.e. at the two-digit level of national industrial classification, whereas the polluting nature of industries is discernible at a more disaggregated level of classification.

The present paper fills the gap in the existing literature, ***firstly*** by distinguishing the polluting nature of industrial agglomeration at different levels of disaggregation of spatial units, and tracing its evolution pattern across Indian districts between the period 2000-01 and 2009-10. ***Secondly***, while estimating the degree of industrial agglomeration economies using the Ellison Glaeser Index, we control for the regional spillover effect (spatial weights).

Thirdly, unlike previous studies, we estimate the plant-level agglomeration economies at a finer industrial classification, defined at the 4-digit level of National Industrial Classification (NIC 2008).

Industrial agglomeration at a disaggregated level can better reflect the polluting[4] nature of industries, and would facilitate policy formulation to mitigate the environmental problems arising from the concentration of specific industries. For example, the industry of leather tanneries versus manufacturing of leather products differs in their polluting nature; these can be distinguished at a four-digit level of industrial classification, whereas at the two-digit level they are clubbed together under manufacturing of leather.

We find that close to half the Indian industries (44%) in Organised manufacturing are highly agglomerated, and most of the agglomerated industries are low-tech and polluting in nature. The overall degree of agglomeration of organised manufacturing sector, however, has decreased over time at both state and district spatial scale; indicating that manufacturing has been spreading across the country into other districts and states.

The industrially advanced states remain the hub of high-tech industries. Low-tech and medium-high-tech manufacturing plants are observed to be spreading in the north-eastern states. Moreover, the second-tier cities witnessed a rise in the share of polluting industries as opposed to the first-tier cities, especially in states like Maharashtra and Rajasthan. This indicates that the cost associated with the increase in the concentration of polluting industries over time has initiated the dispersion process. During the period of analysis, some of the laggard states have gained in terms of the share of polluting industries. This is coupled with a decline in the share of polluting industries in some of the industrially advanced states.

The rest of the paper is Organised as follows: Section 2 briefly describes the evolution of several indices used for empirical estimation of industrial agglomeration economies; Section 3 examines the evolution of the agglomeration process across Indian districts between 2000-01 and 2009-10; and Section 4 concludes the study with some policy suggestions to ensure the sustainable development of industries across India.

Section 1. Indices for empirical estimation of Industrial Agglomeration Economies

With the evolution of theories and understanding of factors underlying industrial agglomeration economies, several indices have been developed to empirically estimate the degree of industrial agglomeration across spatial units (Ellison and Glaeser 1999, Maurel and Sedillot 1999, Guimaraes et al 2011, Amirapu et al 2019). It has been observed that the degree of agglomeration varies both across different levels of spatial units, as well as different levels of industrial classification (i.e. agglomeration of industry measured at the two-digit level differs from the agglomeration measured at the four-digit level) (Maurel and Sedillot 1999, Devereux et al. 2004).

The existing indices can be broadly categorised into two categories -- discrete indices of industrial agglomeration, where spatial units are discrete (Hoover's 1936, Krugman 1991, Ellison and Glaeser

1999, Maurel and Sedillot 1999), and continuous indices, where spatial units are considered to be continuous (Duranton & Overman 2005). The continuous indices are distance-based measures, where kernel density function is estimated using the distance between a given pair of plants. This requires accurate location of a plant, which is often unavailable.

The discrete indices can be further grouped into two broad categories: raw measures of geographical concentration and plant-based measures of industrial agglomeration. The raw measures of the geographic concentration of an industry, namely Hoover's Location quotient (1936) and Krugman's spatial Gini coefficient (1991), capture the disparity in the distribution of regional employment (or output) in an industry relative to regional distribution of overall employment (or employment) in a region (Hoover's 1936, Krugman 1991).

One of the major criticisms of raw measures of industrial agglomeration is that these indices do not consider the within-industry plant structure, which may drive the degree of concentration of an industry. Suppose we have two industries, industry 1 and industry 2. Industry 1 is characterised by many plants all of which are concentrated in one specific region whereas industry 2 is characterised by a single plant. Despite having dissimilar within-industry structures, both the industries will show a similar Gini coefficient. In industry 1, concentration may be driven by the region–specific external economies; however, in industry 2, concentration is solely driven by the plant structure within the industry, i.e. the entire production is concentrated within a plant. This feature makes these indices irrelevant for cross-industry comparisons of the degree of agglomeration.

While constructing an index to measure the degree of spatial concentration of an industry, the main challenge has been to incorporate the randomness involved in the agglomeration process (some industries may be agglomerated spatially just by chance). Ellison and Glaeser (1999) proposed a 'location choice model' for an industry, where the probability of choosing a location by an industry is dependent on the natural advantages of that geographic area (availability of raw materials, water and electricity supply, large consumer markets, network of inter-industry linkages) and externalities arising from the co-location of plants within the industry. They defined agglomeration as the geographic concentration of an industry in excess of the plant-level concentration within the industry. This is also known as industrial localisation index.

Similar to Ellison and Glaeser Index (EG), Maurel and Sedillot (MS) formulated another index to measure the degree of industrial agglomeration. Both the indices measure geographic concentration of an industry after controlling for the effect of within-industry concentration. However, while calculating the degree of agglomeration of an industry, the two indices differ in the way they give weightage to the concentration of overall economic activity in a region. For example, if an industry is located in a highly industrialised area, the MS index takes on high value, whereas if an industry is located in a less industrialised area, then the value of the index is lower. In case of the EG index, there is no such distinction made -- the value is same in both cases.

The Gini, Location Quotient, EG, and MS indices capture the concentration of an industry, as they quantify the variability in employment (or output) of an industry across spatial units relative to the national average. Arbia (2001) argued that these indices did not capture the actual geographical location of a production unit with respect to the other adjacent regions i.e. the spatial correlation between the economic activities of region *i* and the economic activities of neighbouring regions. Moreover, using the spatial unit data defined by boundaries, the degree of industrial concentration is calculated within a pre-defined spatial unit. In the spatial econometrics literature this is also termed the modified area unit problem (MAUP) (Anselin 1988, Arbia 2001, Guimaraes et al 2011).

To account for both the neighbourhood effect as well as to correct the MAUP, indices of industrial concentration i.e. Gini, Location Quotient, EG, or MS can be weighed by using the row-standardised[5] spatial weight matrix. The spatial weights matrix captures the spatial dependence between the units of observations. The weights can be generated using the number of neighbors (*contiguity-based*) or the distance between the adjacent observations (*distance-based*) (Anselin 1988). The spatially-weighted indices capture the degree of '*spatial*' agglomeration of an industry in the true sense.

Section 2 Agglomeration economies and Indian organised manufacturing Sector

2.1 Data

The spatial concentration of organised manufacturing in India has been estimated based on the Annual Survey of Industries (ASI) factory-level database. It covers all factories registered under sections 2(m)(i) and 2(m)(ii) of Factories Act of 1948. A factory[6] is the primary unit of enumeration in the survey process. It is defined as any manufacturing unit with an employment of 10 or more workers using power and those with 20 or more workers not using power. Other than manufacturing units, it also covers all electricity undertakings, engaged in transmission, generation, and distribution of electricity. Moreover, some of the units engaged in services like repairing of motor vehicles, water supply, and cold storage also comes under the purview of the ASI survey. In this study, our analysis is restricted solely to units engaged in the manufacturing process[7].

The sampling frame of the ASI data has undergone several revisions over the years, in order to expand its coverage in each state, as well as across states. The survey frame of ASI can be broadly divided into two categories *viz*, census sector and sample sector. The census sector consists of large plants, based on the number of workers employed. The threshold to define the census sector plants has varied between 50 and 200 workers over the year, so that plants with 200 workers are always surveyed annually.

However, no threshold is followed while sampling plants located in six industrially less-developed states viz, Manipur, Meghalaya, Tripura, Andaman and Nicobar Island, Mizoram, and Sikkim. All the small manufacturing units, not classified under census sector, are included in the sample. The plants defined under the sample sector are randomly surveyed over the period.

The sampling stratum of a manufacturing unit is defined by its geographical location, viz, state and district[8], industry group (at the 4-digit level of NIC), and sector. The multiplier weights are used to generate estimates at these four sub-sample levels i.e. state, district, industry group and sector. The availability of geographical location of a factory, along with the other characteristics like output, raw materials (including types of fuel consumed), types of fixed assets[9] used in the production process, workers employed in each unit, ownership structure and export share, makes this database ideal for analysing the pattern and the underlying agglomerating or dispersing forces in driving the spatial development process of the organised manufacturing industries in India.

Agglomeration is a gradual process, where changes reflect over time. While analysing the evolution of industrial concentration over time in **Section 3**; comparison has been done between the patterns of industrial agglomeration for the period 2000-01 vs. 2009-10 across districts. The degree of concentration of only 111 industries (defined at the four-digit level of NIC2008) could be compared between 2000-01 and 2009-10.

According to the Census of India, between 2001 and 2011, the district boundaries underwent several changes, wherein many new districts were carved out from the existing ones. The number of districts increased from 593 in 2001 to 640 in 2011. This change is also reflected in the coverage of the ASI unit-level database. The ASI 2000-01 round covered 455 districts, as opposed to ASI 2009-10 round, wherein the coverage increased to 593 districts. While comparing the evolution of concentration of polluting industries across districts between ASI round 2000-01 and ASI round 2009-10, the new districts were mapped with the earlier amalgamated districts. The detailed mapping is summarised in the appendix **Table A.1.**

The agglomeration economies have been estimated using the Ellison Glaeser Index; based on the plant-level employment data,[10] defined at the four-digit level of NIC-2008. It captures the labour pool effect, i.e. the externalities arising from the sharing of labour between plants within the same industry. The paper also uses spatially-weighted Ellison Glaeser Index to account for the regional spillover or *neighbourhood effect* of industrial agglomeration economies.

While estimating the neighbourhood effect, we utilised the Shapefiles of Indian states and districts, as published by the Indian Institute of Remote Sensing (IIRS). The shapefiles are vector files containing geo-spatial information, including the latitude and longitude of each district and state. The information on coordinates was used to calculate the spatial-weight matrix using the GeoDa Software[11]. The detailed technique of spatially weighing the EG index has been explained in the **Appendix** in **Section A.1.**

While analysing the nature of industrial agglomeration in India's manufacturing sector, we have categorised the industries in terms of their technology intensity and polluting nature. The OECD definition of technology intensity[12] of industries has been followed (OECD 2011). The industries are classified into four major groups: Low-tech, Medium low-tech, Medium-High tech and High-tech.

The detail mapping of NIC codes into OECD classification has been mentioned in the **Appendix, Table A.2.**

The CPCB of the Government of India has classified industries into four different categories (Red, Orange, Green, and White) based on the pollution index score of each industry[13]. The pollution index score is dependent on the four criteria:

i. emission from the industries (air pollutants),

ii. effluents from industries (water pollutants),

iii. hazardous wastes generated by industries, and

iv. consumption of resources by industries (CPCB 2016).

In this study, the red and orange category industries are defined as *polluting industry*; Green and White category industries have been defined as *non-polluting industries*. This categorization was initiated by CPCB to regulate the location of some highly-polluting industries in ecologically sensitive areas across Indian states,[14] and curb operations of certain high-pollution industrial processes.

2.2 Estimation of Industrial Agglomeration Economies

We have used the spatially-weighted Ellison Glaeser Index[15] (γ_0^{sw}) to estimate the degree of agglomeration of organised manufacturing industries; defined at the 4-digit level of National Industrial Classification, across Indian districts. In the year 2009-10, the average degree of agglomeration of the Organised manufacturing industries is found to be moderate, with an EG^{SW} of 0.047.

While estimating the degree of agglomeration of 130 Organised manufacturing industries, defined at the 4-digit level of industrial classification, 48% of the industries were found to be highly agglomerated (i.e. $\gamma_0^{sw}>0.05$), 33% were moderately agglomerated (i.e. $0.05<\gamma_0^{sw}\leq 0.02$), and 19% were found to have a lower degree of agglomeration (i.e. $\gamma_0^{sw}<0.02$) across Indian districts in the year 2009-10.

Most of the low-tech and medium-low-tech industries are found to be highly agglomerated across Indian districts. 58% of the highly agglomerated industries are non-polluting, and 42% of the highly agglomerated industries are polluting in nature. The top 10 polluting industries which are found to be highly agglomerated across Indian districts in the year 2009-10 have been listed below in **Table 1**

Table 1. Highly agglomerated polluting industries

Industry Description (NIC)	Technology Intensity	Degree of Agglomeration (γ_0^{sw} >0.05)
Manufacture of knitted and crocheted fabrics (1391)	Low tech	0.442
Manufacture of knitted and crocheted apparel (1430)	Low tech	0.430
Manufacture of tobacco products (1200)	Low tech	0.264
Manufacture of other chemical products n.e.c. (2029)	Low tech	0.216
Tanning and dressing of leather; dressing and dyeing of fur (1511)	Low tech	0.203
Manufacture of other textiles n.e.c. (1399)	Low tech	0.159
Manufacture of articles of fur (1420)	Low tech	0.140
Finishing of textiles (1313)	Low tech	0.129
Manufacture of carpets and rugs (1393)	Low tech	0.118
Processing and preserving of fish, crustaceans and molluscs and products thereof (1020)	Low tech	0.099

Source: Author's calculation based on ASI unit level database

Section 3: Evolution of Industrial Agglomeration

3.1 Overall degree of agglomeration of Organised manufacturing industries has declined during the period 2000-01 to 2009-10.

We find that the overall degree of manufacturing agglomeration has been declining during the period 2000-01 to 2009-10 across Indian districts. This is in line with the phenomenon observed by Ghani et al (2012) and Colmer (2014), who noted that the share of the Organised manufacturing sector in urban areas has declined over time. The degree of manufacturing agglomeration has been moderate (γ_0 ≤0.05) during the period 2000-01 to 2009-10.

During the period of our analysis, the manufacturing output of the Organised sector has grown by 15%, accompanied by a sluggish employment growth of 4% in the sector. While analysing the distribution of manufacturing activity across Indian states, the share of the already-industrialised states like Gujarat, Maharashtra, Andhra Pradesh, and Uttar Pradesh constitutes the bulk of the total manufacturing output. States like Maharashtra, Tamil Nadu, Uttaranchal, Gujarat, and Uttar Pradesh have attracted the maximum number of new plants[16] during the year 2009-10. However, during this period it is noteworthy that some of the industrially laggard states, like Jammu and

Kashmir, Meghalaya, Uttarakhand, and Himachal Pradesh, have registered a high growth rate in terms of both manufacturing output as well as employment. This suggests that the congestion costs associated with the agglomeration process act as dispersing forces, leading to the spread of industries across regions.

3.2 Medium-high tech industries are dispersing across districts as opposed to High-tech industries

While analysing the nature of the evolution of industries across space over time, plants belonging to low-tech and medium-high-tech industries are spreading more across the districts as opposed to high-tech plants. **Figure 1** (below) depicts the distribution of manufacturing plants in *high-tech* industries across Indian districts. High-tech plants are concentrated in a few districts in Gujarat, Maharashtra, West Bengal, Karnataka, Tamil Nadu, Andhra Pradesh, Delhi, and Chandigarh. We observed that high-tech plants are spreading out across districts within the same states. Districts like Rann of Kutch of Gujarat, or Faridabad and Gurgaon of Haryana, show a significant increase in the percentage share of high-tech plants in the year 2009-10 as compared to the year 2000-01.

Figure 1 Distribution of High-tech Plants across Indian District 2000-01 vs. 2009-10

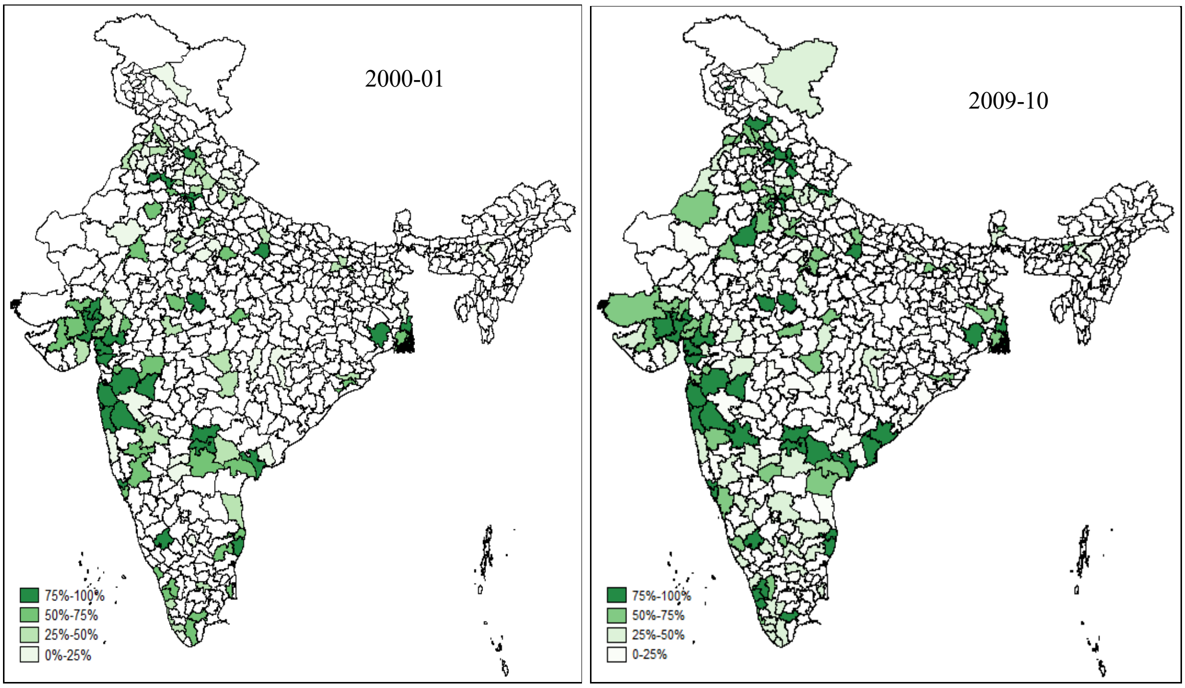

Fig 1a) Distribution of High-tech Plants in 2000-01 Fig 1b) Distribution of High-tech Plants in 2009-10

In Figure 2, the distribution of medium-high-tech plants is seen to be spreading away from the coastal belt to the central and northern districts of India. Some districts in the interior states of Haryana, Punjab, and Delhi show a significant increase in the percentage share of medium-high-tech plants. Within Maharashtra, medium-high-tech industries seem to have spread to districts like

Aurangabad, Satara, and Nashik. Some districts of Karnataka and Goa also seem to have gained in terms of these plants. On the western side, many districts of Rajasthan like Bikaner, Ajmer, Jodhpur, and Jaipur seemed to show a significant rise in the number of medium-high-tech plants in the year 2009-10 as compared to the year 2000-01.

Figure 3, reflects that plants in low-tech manufacturing industries are spreading out across the country over the entire period of our analysis. In contrast to high-tech and medium-high-tech, low-tech industries have also spread to the north-eastern districts of India, which are the least industrialised regions of the country.

Figure 2 Distribution of Medium-High tech Plants across Indian Districts 2001 vs. 2010

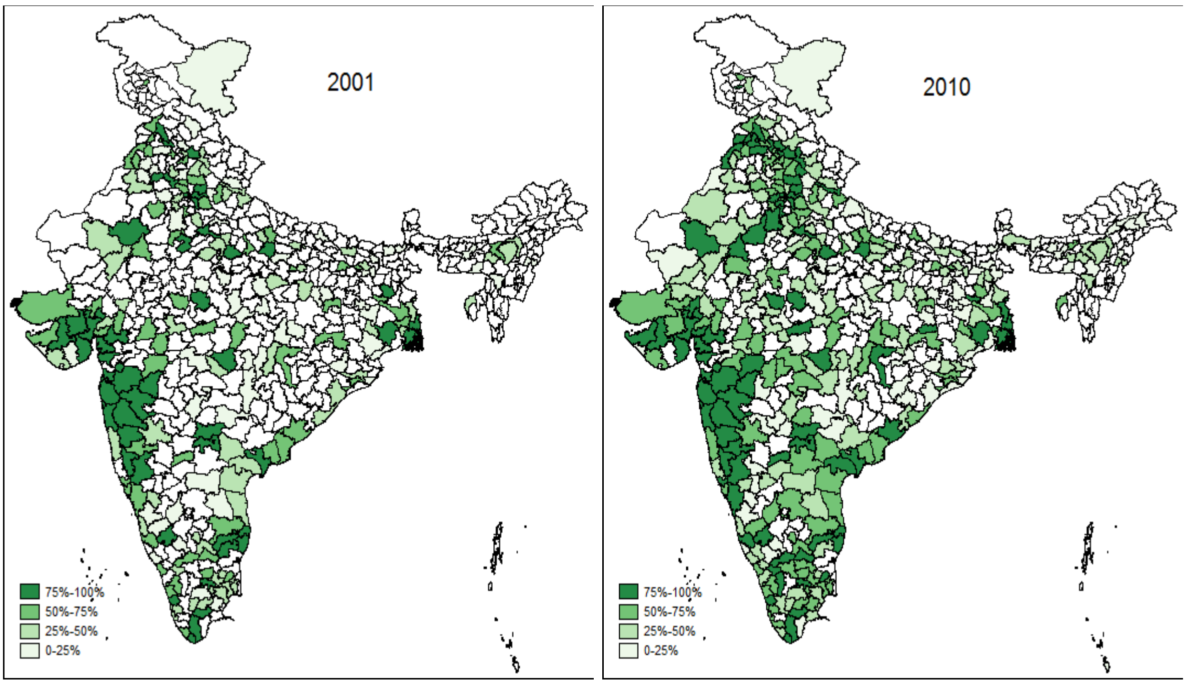

Fig 2a) Medium-High-tech Plants in 2001 Fig 2b) Medium-High-tech Plants in 2010

From the preliminary analysis of the distribution of high-tech, medium-high-tech, and low-tech plants across districts, we can conclude that high-tech manufacturing industries are biased towards high-income states like Gujarat, Maharashtra, Tamil Nadu, and Andhra Pradesh. However, plants belonging to medium-high-tech industries have spread, over time, to northern districts like Faridabad, Ghaziabad, and Gurgaon in Uttar Pradesh and Haryana, which have become hubs of the automobile industry. The low-tech industries show a more even distribution across states, including low-income, least-industrialised North-Eastern states.

Figure 3 Distributions of Low-tech Plants across Indian Districts 2001 vs. 2010

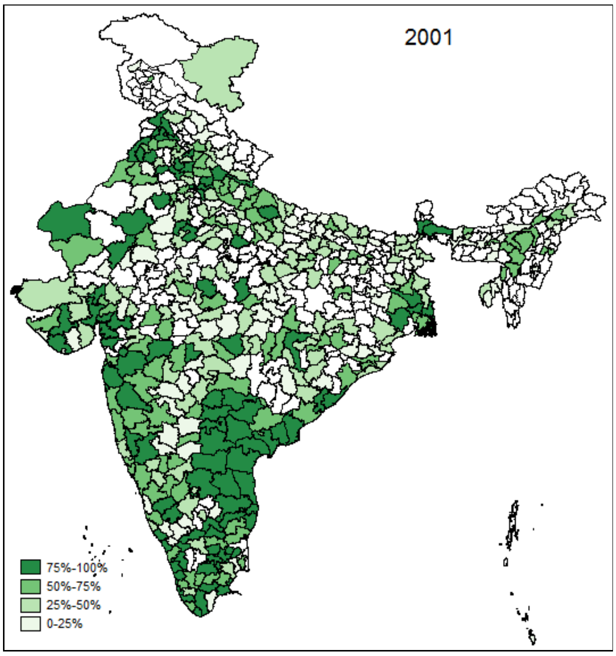

Fig 3a) Distribution of low-tech Plants in 2001

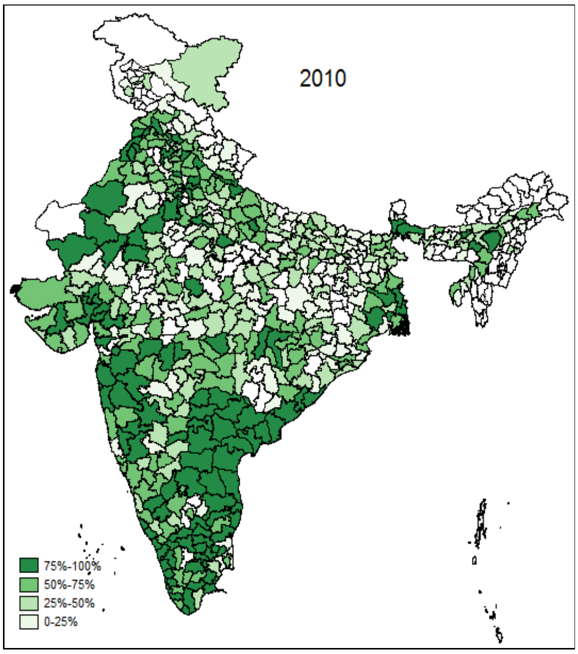

Fig 3b) Distribution of low-tech Plants in 2010

3.3 Second-tier cities have registered growth in the number of polluting industries

While distinguishing between polluting and non-polluting industries, it has been observed that during 2000-01 to 2009-10, the dispersion of polluting industries has been higher as opposed to the non-polluting industries[17]. **Figure 4** (below) indicates the distribution of plants across districts belonging to polluting industries in the year 2000-01 vs. 2009-10. Changes in the distribution of polluting industries can be observed in small pockets, as highlighted by the yellow dotted circles.

Some of the districts of the coastal states in the south have become cleaner, in terms of reduction in the number of polluting plants. Some of the districts of northern states and north-eastern states are observed to have seen a rising share of polluting plants over this time. During this period, some of the second-tier cities like Faridabad, Coimbatore, Ludhiana, Pune, Jaipur, and Rajkot have registered an increase in the number of polluting plants.

The spurt in the process of urbanization in Tier-1 cities has led to a massive outflow of people toward second-tier cities. Moreover, it has been established in the literature that over the last few years, the rising cost of land and other factors of production has led to the shift of the manufacturing industries towards semi-urban areas or rural areas (Colmer 2014). Since manufacturing is associated with emissions, the concentration of industries (especially polluting industries) within metropolitan cities will aggravate the cost of congestion in terms of environmental pollution. This in turn may have further fuelled the process of the shift of manufacturing industries toward the secondary cities.

Figure 4: Distribution of Plants across Districts

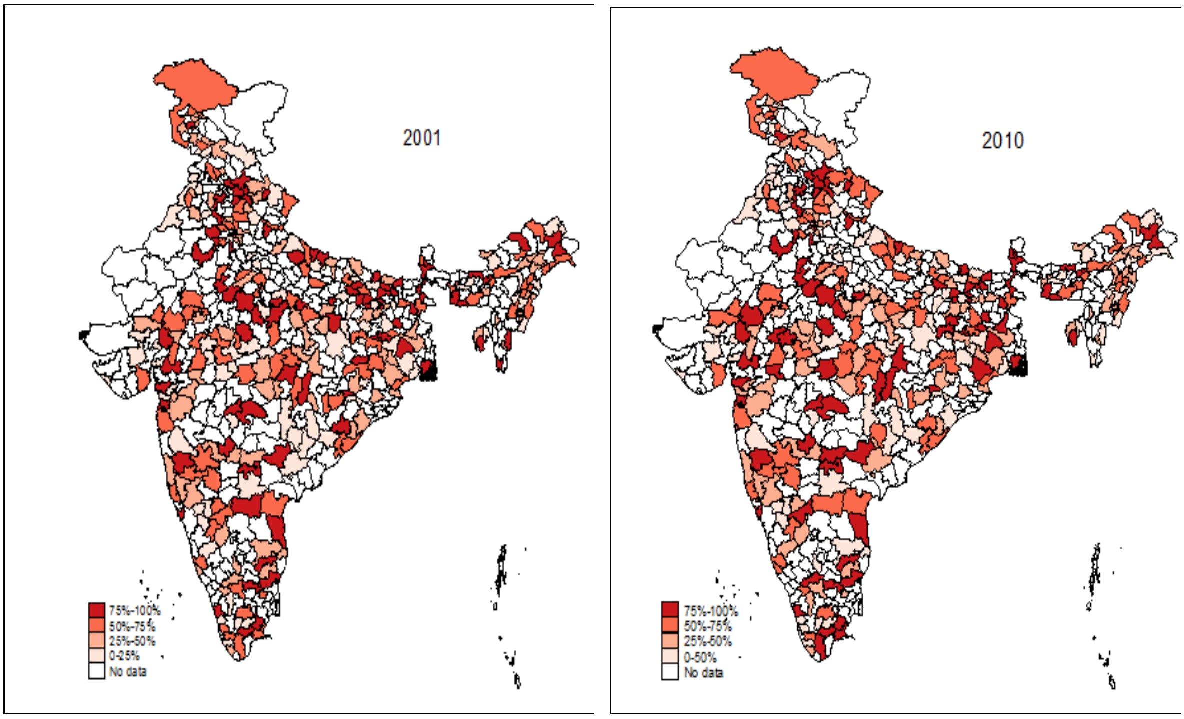

Fig 4a) Distribution of Polluting Plants in 2001 Fig 4b) Distribution of Polluting Plants in 2010

Section 4: Conclusion

This paper tracked the evolution pattern and nature of industrial agglomeration in the Indian Organised manufacturing sector (at a 4-digit level of industrial classification across districts. The agglomeration economies estimated in this paper capture the positive externalities arising due to sharing of the labour pool across plants, reflecting the employment-generation potential of industrial clustering policies.

We find that the Organised manufacturing industries seem to be moderately agglomerated, and over time, showed a decline in the degree of agglomeration at the district level. This indicates the dispersion process of Organised manufacturing activities across districts over time.

While analysing the nature of agglomeration, the low-tech and medium-low-tech industries are found to be highly agglomerated. Some of the highly agglomerated low-tech industries are also polluting in nature. This raises concern about the environmental impacts of the clustering process of manufacturing industries. The diseconomies associated with the process threaten the sustainability of the cluster development programmes, thereby weakening the objective of achieving manufacturing industry-based economic growth and employment generation. Analysing the nature (polluting nature as well as technology-based nature) of the industrial clusters seems to be an inevitable part of the manufacturing-based cluster development programmes.

We also find that medium-high-tech industries and low-tech industries have dispersed across districts, with the latter group spreading to poorer regions in northeast India. Most of the high-tech plants are found to agglomerate in the districts of high-income states. The technological difference may be one of the factors behind fostering regional income inequality across Indian districts. This needs further empirical analysis and can be a future research area.

During the period of analysis, the polluting industries are also found to be dispersing across Indian districts. It appears that, over time, some of the districts of the industrially laggard states have gained a higher share of plants belonging to polluting industries. This is coupled with the observation that some of the high-income districts have seen a drop in the share of plants belonging to polluting industries. It will be interesting to examine the role of (the degree of) environmental stringency in driving this dispersion process, after controlling for other agglomeration externalities. The testing of the 'pollution haven effect' across Indian districts in presence of agglomeration externalities can similarly be a potential future area of research.

The paper also observes that the spillover of economies (or diseconomies) across Indian districts cannot be limited by pre-defined district boundaries. While estimating the degree of agglomeration of industries, it is imperative to consider the degree of agglomeration of industries in adjacent regions. A recent initiative of the Government of India -- to develop a Comprehensive Zoning Atlas at the district level for new investors, which incorporates both economic as well as environmental parameters -- may ensure the long-term sustainability of cluster development programmes in India.

References

Barua, A., & Sawhney, A. (2015). Development policy implications for growth and regional inequality in a small open economy: The Indian case. *Review of Development Economics*, *19*(3), 695-709.

Colmer, J. (2015). Urbanization, growth, and development: evidence from India. *Unpublished review paper*.

Das, Samarjit, Ghate, Chetan & Robertson, Peter E. (2015). Remoteness, Urbanization, and India's Unbalanced Growth. World Development, 66, 572-587.

Devereux, M. P., Griffith, R., & Simpson, H. (2004). The geographic distribution of production activity in the UK. Regional Science and Urban Economics, 34(5), 533-564.

Ellison, G., & Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? The American Economic Review, 89(2), 311-316.

Ghani, E., Goswami, A. G., & Kerr, W. R. (2012). Is India's manufacturing sector moving away from cities? National Bureau of Economic Research Working Paper No. w17992, Washington D.C.

Ghani, E., Kerr, W. R., & O'Connell, S. (2014). Spatial determinants of entrepreneurship in

Guimaraes, P., Figueiredo, O., & Woodward, D. (2011). Accounting for neighboring effects in measures of spatial concentration. Journal of Regional Science, 51(4), 678-693.

Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economic Statistics*, 162-171.

India. Central Pollution Control Board. (2009, 2011, 2013). *Comprehensive environmental assessment of industrial clusters*. Central Pollution Control Board.

India. Regional Studies, 48(6), 1071-1089

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483-499.

Lall, S. V., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. Journal of Development Economics, 73(2), 643-673.

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

Mukim, M. (2014) Coagglomeration of formal and informal industry: evidence from India. Journal of Economic Geography, 15(2), 329-351.

APPENDIX

Table A.1 Mapping of New Districts@

State	Name of the New Districts (2009-10)	Earlier Amalgamated Districts 2000-01
Punjab	Sahibzada Ajit Singh Nagar	Rupnagar and Patiala
Haryana	Mewat	Gurugaon
Bihar	Arwal	Jehanabad
West Bengal	East Medinipur	Medinipur (West Medinipur)
Jharkhand	Simdega	Gumla
	Jamtara	Dumka
	Saraikela-Kharsawan	Paschim Singhbhum
Madhya Pradesh	Anuppur	Shahdol
	Burhanpur	East Nimar
Tamil Nadu	Krishnagiri	Dharampuri

@ Mapping of districts only covered under the ASI 2000-01 round and ASI 2009-10 rounds has been tabulated above

Section A.1

Ellison-Glaeser industrial agglomeration index

While estimating the degree of agglomeration of an industry, Ellison and Glaeser (1999), constructed a discrete probability model (following Bernoulli distribution) to analyse the correlation between the location choices of two plants belonging to the same industry. The two plants within the same industry may locate near each other due to the presence of externalities or spillovers. The benefits from locating near other plants arises from exchange of labour pool or of technological know-how within the same industry, or inter-plant trade in intermediate inputs.

In our analysis, we focus on the labour pool channel. Since a plant chooses to locate in a region where it can gain maximum profit, the profit function of a plant belonging to industry i located in region m is affected by two factors: a) employment share of region m in aggregate employment, and b) location of other plants within the same industry owing to the presence of spillovers.

Let there be N number of plants in industry i and $q_1, \ldots q_j, \ldots, q_N$, are the shares of these plants in the total employment (or output) of the industry. The Herfindahl index of industry i $H_i = \sum_{j=1}^{N} q_j^2$, captures the plant size distribution within industry i.

The model assumes that the location choice of plant j to set up its operations is an independent identically distributed random variable, so the regional share of industry i can be re-written as, $s_i = \sum_{j=1}^{N} q_j u_m$, where u_m is the Bernoulli random variable (which takes a value of 1 if a plant j locates in region m, and 0 otherwise).

Ellison and Glaeser modelled the interaction between the location decision of two plants *j* and *k* within the same industry *i* owing to the presence of spillovers. The interaction between the location decisions of two plants within the same industry in region *m* is defined as,

$$Corr(u_{mj}, u_{mk}) = \gamma_0 \quad for\ j \neq k \qquad (1a)$$

Where γ_0 captures the degree or the strength of spillover between two plants belonging to the same industry, located in the same region. The probability that plant *j* and *k* will locate in the same area *m* is given by,

$$p(j_m k_m) = E(u_{jm} u_{km}) = Cov(u_{jm}, u_{km}) + E(u_{jm})E(u_{km}) = \gamma_0 x_m(1 - x_m) + x_m^2$$

The probability P that plant *j* and *k* locate in any of the *M* locations is given by

$$P = \sum_{m=1}^{M} p(j_m k_m) = \gamma_0 x_m(1 - x_m) + x_m^2$$

$$P = \sum_{m=1}^{M} p(j_m, k_m) = \gamma_0 (1 - \sum_{m=1}^{M} x^2{}_m) + \sum_{m=1}^{M} x_m{}^2 \qquad \text{(ii)}$$

Ellison and Glaeser explained (using the example of throwing a dart in space) that the location choice of a plant is a two-stage process. In the first stage, natural advantages of a region drives a fraction of the plants to locate there. In the second stage, some plants choose to co-locate in the same region owing to the presence of spillover among them. The strength of the spillover is captured by parameter γ_0.

$$\gamma_0 = EG_i = \frac{G_i - (1 - \sum_{m=1}^{M} x_m^2) * H_i}{(1 - \sum_{m=1}^{M} x_m^2) * (1 - H_i)} \qquad \text{(iii)}$$

where, G_i is the measure of raw concentration of the industry as defined by equation (2)

Spatially-Weighted Index of Industrial Agglomeration

The Ellison-Glaeser index has been criticised for capturing the degree of concentration irrespective of its geographical position relative to other areas within the country, i.e. spillover from adjacent regions was not considered in the estimation process.

Guimareas et al. (2010) index of industrial agglomeration has been modified to incorporate the spillover effect of the neighbouring regions. The spillover effect of economic activity of adjacent regions has been captured by weighing the regional share in equation (iii) by using the spatial weights matrix, W.

The modified Ellison-Glaeser Index of agglomeration can thus be re-written as

$$\gamma^{SW}{}_{i} = \frac{G_s - (1 - W\sum_{m=1}^{M} x_m^2) * H_i}{(1 - W\sum_{m=1}^{M} x_m^2) * (1 - H_i)} \qquad \text{(iv)}$$

Since, W is a spatial-weight matrix, the equation (iv) can be re-written in the vector form as

$$\gamma^{SW}{}_{i} = \frac{G_s - (1 - x_m{}' W x_m) H_i}{(1 - x_m{}' W x_m)\sum_{m=1}^{M}(1 - H_i)} \qquad \text{(v)}$$

where, $G_s = (s_{im} - x_m)' W (s_{im} - x_m)$ is the Spatially weighted Gini Index of equation (2)

Table A.2 Technology classification of industry

ISIC Rev.4/ NIC 2008	Industry Description	OECD tech classification
10	Manufacture of Food Products	Low
11	Manufacture of Beverages	Low
12	Manufacture of Tobacco Products	Low
13	Manufacture of Textiles	Low
14	Manufacture of wearing apparel	Low
15	Manufacture of leather and related products	Low
16	Manufacture of Wood and Wood Products	Low
17	Manufacture of Paper and Paper Products	Low
18	Printing and Reproduction of Recorded Media	Low
19	Manufacture of Coke and Petroleum Products	Medium-low
20	Manufacture of chemicals and chemical products	Medium-high
21	Manufacture of basic pharmaceutical products & preparations	High
22	Manufacture of Rubber and Plastic Products	Medium-low
23	Manufacture of Other Non-Metallic Mineral Products	Medium-low
24	Manufacture of Basic Iron and Steel	Medium-low
25	Manufacture of Fabricated Metal Products	Medium-low
26	Manufacture of computer, electronic and optical products	High
27	Manufacture of electrical equipment	Medium-high
28	Manufacture of machinery and equipment	Medium-high
29	Manufacture of motor vehicles, trailers and semi-trailers	Medium-high
30*	Manufacture of other transport equipment: aircraft & spacecraft	High
30*	Manufacture of other transport equipment - railway equipment	Medium-high
301	Building of ships and boats	Medium-low

* excluding 301 - Building of ships and boats

Source: UNIDO-World Bank definition based on OECD classification

NOTES

[1] Calculation based on data from World Development Indicator database.
[2] The CEPI score has a scale of 0-100, where a score of 70 and above indicates the region is critically polluted, and a score of 60-70 indicates a severely polluted area. The criteria pollutants include (i) Sulphur dioxide, nitrogen dioxide, fine particulate matter (PM_{10}, $PM_{2.5}$), lead, ozone, carbon monoxide, benzene, etc for ambient air quality; and (ii) Dissolved oxygen, biological oxygen demand, chemical oxygen demand, pH, faecal coliform, phosphorous, ammonia, arsenic, heavy metals including lead, cadmium, mercury, etc for surface and groundwater quality. The calibration of CEPI was revised in 2016, so the index scores are not quite comparable with the earlier years.
[3]http://www.cpcbenvis.nic.in/cpcb_newsletter/ZONING%20ATLAS%20FOR%20SITING%20OF%20INDUSTRIES.pdf
[4] While categorizing an industry as polluting or not polluting, we follow the Red list, Orange list, and Green list of industries as defined by the Ministry of Environment, Forests, and Climate Change. The categorization is based on the emission/effluent load of the industries and consumption of resources.
[5] Each element in the row of the spatial weight matrix is standardised by the row total. This is a standard exercise in spatial econometrics literature to assign equal weightage to all the neighbors of a particular spatial unit.
[6] The owner of each factory identified under some industry group has to file a return annually to the statistical office of the regional offices of NSSO. However, owners with more than two factories identified under the same industry group and located in the same state are allowed to file consolidated or joint returns.
[7] According to the NIC2008 classification, all units categorised under divisions 10 to 32 are included in the study.
[8] The information on the district codes has been suppressed beyond the year 2009-10 owing to the confidentiality issue.
[9] Book value of fixed assets is reported in the Annual Survey of Industries data.
[10] Alternatively, as a robustness check, output data has also been used to estimate the industrial agglomeration economies.
[11] GeoDa is a free and open-source software tool and is widely used for spatial data analysis. The software has special features for spatial data modelling *viz* calculation of spatial weight matrix, spatial auto correlation statistics, spatial regression analysis, Moran-I statistics, etc.
[12] The definition of technology intensity of industries is based on the expenditure on research and development.
[13] The Red category is defined as industries with a pollution index score >60; the Orange category is defined as industries with a pollution index score greater than or equal to 41 but less than 60; the

Green category is defined as industries with a pollution index score greater than equal to 21 but less than 41; the White category is defined as industries with pollution index score less than equal to 20.
[14] The ecologically sensitive areas are protected areas for the conservation of Biodiversity; for example, Doon Valley in Uttarakhand, and Sultanpur in Uttar Pradesh.
[15] EG index (γ_0) captures the degree or the strength of spillover between two plants belonging to the same industry, located in the same region. Further, spatially weighted EG also corrects for the spillover effect of the adjacent regions. the detail of the index has been illustrated in the **Appendix**. The value of the index ranges between -1 <γ_0 <1. If the value of the index is 0 then it indicates a *lack of agglomerative forces*. If the value of the index for an industry is greater than 0 then the industry is localised. The thresholds to classify different industries- If the value is below 0.02 but positive then the industry is *not very agglomerated*. If the value varies between 0.02 and 0.05 then the industry is *moderately agglomerated* and if the value of is above 0.05 then the industry can be categorised as *highly agglomerated*. The negative value indicates that the industry is *dispersed.*
[16] We define new plants an age of less than equal to three years i.e., their year of incorporation is between 2007-08 and 2009-10
[17] Dispersion is also visible in case of some non-polluting industries. However, there are also some non-polluting industries that have become agglomerated over time.