Geographic Concentration in Indian Manufacturing and Service Industries: Evidence from 1998 – 2013*

Amrit Amirapu† Rana Hasan‡ Yi Jiang‡ Alex Klein†

June, 2018

Abstract

This paper uses a comprehensive new data source to document basic facts about geographic concentration among industries in India from 1998 to 2013. Unlike previous studies, our data allow us to accurately measure industrial concentration at the district level and cover manufacturing and services, as well as the formal and informal sectors. Our most striking finding is that average levels of industrial concentration fell dramatically between 1998 and 2013, driven by steep reductions in capital-intensive manufacturing industries. We provide suggestive evidence that this increasing dispersion may be due to improvements in interregional transportation coupled with inefficient land management policies and limited labor mobility.

Key words: Industrial agglomeration, regional inequality, land management, labor mobility, India

JEL codes: O14, O18, R11, R12

Accepted: 27 June 2018

* We are grateful to the editors and referees of the Asian Economic Policy Review. We would particularly like to thank Colin McKenzie, Marcus Noland, Shuji Uchikawa, Rakesh Mohan and other participants at the 27th AEPR Conference in Tokyo for a number of helpful suggestions that have improved the paper. Finally, we wish to thank Radine Michelle Rafols and Marjorie V. Remolador for their support in putting together the data. All errors are our own.
† School of Economics, University of Kent.
‡ Asian Development Bank.

Disclaimer

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of ADB or its Board of Governors or the governments they represent. ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use.

1 Introduction

The spatial distribution of economic activity within countries has profound implications for both efficiency and equity. Understanding where firms choose to locate and why is therefore of considerable policy importance. 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 that witnessed significant investments in transportation infrastructure and 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 during which the Indian economy grew rapidly - but unevenly - so that regional inequalities increased, with some regions forging ahead while others lagged behind.

There is a body of research which tries to understand this process (for example, 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 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) and Guimarães et al. (2011). The paper's main contributions flow from the data that we marshal for the task: establishment-level data from the last three rounds of the Economic Censuses of India.

Economic Census data have not yet been used to study questions of industrial agglomeration - although they are uniquely suited for it. First, they allow us to quantify agglomeration at a suitably geographically disaggregated level of analysis, namely at the district level.[1] Previous studies have relied on sample surveys that are only representative at very large geographic units – usually states, which in India are

[1] Districts constitute the second tier of subnational administration in India.

often the size of large countries, and may thus be inadequate for capturing agglomeration externalities at the relevant scale. Second, most previous studies have focused their analysis exclusively on the formal manufacturing sector. In contrast, our data allow us to characterize industrial concentration in the entire economy (excluding agriculture), including manufacturing and services, and the formal and informal sectors.

This last point is of particular importance as it is the service sector which has been driving India's fast growth rates, despite the fact that most of the reforms of the early 1990s were focused on unleashing India's manufacturing sector from 1) restrictive trade and foreign direct investment (FDI) policies, and 2) 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 EGI 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 EGI 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 geographic concentration among industries in India at the turn of the twenty-first century and the first decade of the new millennium.

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 and other developed countries, for which 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 (for example, textiles). This feature of India's economic landscape may be related to India's history of promoting and directing 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 (for example, 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 manufacturing industries. This trend is in stark contrast with that of China over a similar period (Lu & Tao 2009), where agglomeration levels seem to have increased over time.

These findings have significant implications for policy. A large body of empirical evidence in the literature at large emphasizes the importance of agglomeration as a contributor to productivity growth (for example, Ciccone and Hall, 1996). The fact that industrial activity is becoming more dispersed suggests that India may be missing out on a potentially important source of productivity growth. On the other hand, greater dispersion of industry may help keep spatial inequalities in check. It is thus important to understand what aspects of the policy and broader economic environment may be driving our results. Conclusively answering this question is beyond the scope of this paper, but we close our analysis with some evidence that is consistent with the hypothesis that the observed de-agglomeration has been driven by reductions in trade costs due to improvements in regional transportation

infrastructure in combination with limited labor mobility (in comparison to China, for example) and growing scarcity of affordable land around existing centers of production due to inefficient land management policies – which we term "congestion effects".

The rest of the paper is organized as follows. First, we summarize the related literature and describe our data in detail. Following this we explain the methodology we use to measure agglomeration and coagglomeration, and then introduce a simple interpretative framework based on established theory from the agglomeration literature which will be used to discuss the results. With the interpretive framework in place, we go over the results of our analysis and conclude with a discussion of the policy implications of the 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) examine the geographic concentration of French manufacturing industries in 1993 and finds similarities between France and the US regarding the most and least localized industries. Maurel and Sedillot 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. Their evidence shows that the extent of geographic concentration in China's manufacturing industries, as measured by the Ellison-Glaeser index (EGI), has increased steadily between 1998 and 2005. The trend appears to reflect the effects of the economic reforms that China

launched in 1978, whereby market forces were increasingly relied upon in resource allocation. However, comparing the degree of agglomeration in 2005 with that of developed countries in the late 1980s and 1990s, Lu and Tao find that China's manufacturing industries still had a significantly lower level of agglomeration, which they attribute to protectionism at local levels (for example, across provinces) .

Turning to India, there are a number of studies that examine various aspects of agglomeration of economic activity. The closest to ours in terms of the methodology used to quantify agglomeration is by Fernandes and Sharma (2012). They examine the geographic concentration of Indian manufacturing between 1980 and 1999 using plant level data on formal sector firms provided in the Annual Survey of Industries (ASI).[2] Fernandes and Sharma find that the average EGI was stable in the 1980s and decreased considerably in the 1990s. Furthermore, they provide evidence that de-licensing and FDI liberalization led to reduced spatial concentration of manufacturing, whereas trade reforms had no significant effect.

One reason for this link offered by Fernandes and Sharma is that India's industrial licensing regime inadvertently created inefficient manufacturing clusters; reform of the regime has thus worked to disperse these clusters. This raises the question of what might lie behind the inefficiency of incumbent clusters. Some clues are to be found in the work of Lall et al. (2005), who also rely on plant level data on formal manufacturing firms and compute Gini coefficients for the distribution of sector-specific employment shares of 11 manufacturing sectors across districts in 1994-95. They find these to be relatively high, indicating that industries tended to concentrate in a relatively small number of districts. Turning to the potential drivers of this concentration, Lall et al. estimate a production function that includes various potential sources of agglomeration economies as inputs to production and find that access to markets is productivity enhancing; in contrast, locating in dense urban areas does not provide any productivity benefits (and often adds to costs). They offer two reasons why the spatial concentration of industries is high despite weak benefits from locating in dense urban areas. First, regulatory barriers that prevent Indian firms from closing plants (such as those imposed by labor regulations). Second,

[2] Formal sector firms in the industrial sector are those employing 10 or more workers (20 or more for firms not using electricity) and that are registered under the Factories Act.

weaknesses in the transport infrastructure linking smaller urban areas to the interregional transport network; this would prevent manufacturing activities, at least standardized ones, moving from large and costly agglomerations to lower cost, secondary centers of agglomeration. Notably, the Indian government made large investments in national transportation infrastructure in the period after that studied in Lall et al. (2005), which we will have more to say about in our own analysis.

The possibility that something may have prevented firms from moving to secondary centers is also a theme of the recent work by Desmet et al. (2015), who examine the relationship between district level employment growth in an industry and the density of employment in that district. Desmet et al. find that manufacturing employment has grown faster in districts with lower initial manufacturing employment density between 2000 and 2005, thus implying greater dispersion. In contrast, districts with low and high service employment density have experienced faster growth of employment in services as compared to medium service density districts between 2000 and 2006.[3]

Our study contributes to this literature in several ways. First, as we discuss in more detail below, we use data from India's Economic Census, which allows us to cover comprehensively not just manufacturing, but also services.[4] The other significant advantage of the Economic Census is that it is representative at divisions more disaggregated than the State, unlike the other establishment-level datasets relied upon in the previous literature. Second, we cover a time period which extends well beyond the 1990s, when India undertook a series of liberalizing reforms. Finally, we undertake a closer look at the forces of dispersion that seem to underlie the patterns and trends characterized by the existing literature.

3 Data

[3] 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 manufacturing became less urbanized.

[4] The previous studies, even when they include services, rely on survey data that cover only particular parts of the service sector.

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.[5] 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 the district and town/village level.[6] 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 most 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 the manufacturing and service sectors.

The EC 1998, EC 2005 and EC 2013 are based on the household listings and district definitions of the Population Census for 1991, 2001 and 2011, and adopt the National Industrial Classification of 1987, 2004 and 2008, respectively. In order to account for changes in district boundaries and industrial

[5] We use the terms establishment, enterprise, and firm interchangeably.
[6] A village is the rural counterpart to a town.

definitions across rounds, we matched the administrative boundaries of districts and industry classifications across the three ECs. 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[7] 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).

Table 1 presents counts of establishments and employment by EC and sector in India from 1998 to 2013. The number of non-agricultural establishments increased from 26.9 million in 1998 to 45.4 million in 2013. Total employment in the non-agricultural industries increased at an annual rate of 2.3% 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 the 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 National Sample Survey Organisation's (NSSO's) Employment and Unemployment Survey (EUS). 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 (Amirapu & Gechter, forthcoming). This is likely to be due to under-coverage of the very smallest firms in the Economic Census (mostly own account enterprises).[8] With this caveat, the Economic Census appears to be a good data 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

[7] There are 59 in the 1998 and 2005 EC waves. Three industries are not found in the 2013 EC due to changes in industry categorizations.
[8] This is also the conclusion of Manna (2010) and Unni and Raveendran (2006).

establishments with 5 or more workers. The summary statistics for such establishments are also presented in Table 1.

[Table 1 around here]

4 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.[9]

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 $S'=[s_1, s_2, \ldots, s_M]$ 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 component G_i is at the heart of the index and captures the extent to which a particular industry has a higher concentration of employment located in certain districts, relative to the total employment of those districts. 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. Larger values of the EG index imply greater geographic concentration, controlling for the size distribution. Ellison and Glaeser (1997) consider industries to be strongly concentrated if

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

they have an EG index value of 0.05 or above, and weakly concentrated for values below 0.02. An EG index value of zero implies that plants are located perfectly randomly.

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, 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, G_i, and Ψ is a spatial weight matrix. Ψ is defined as $\Psi = W + I$ where I is the identity matrix and 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 $\Psi = I$ (that is, 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.

5 Interpretative Framework

Before presenting our results we introduce an interpretative framework – based on theory from the agglomeration literature and what tends to be called 'new economic geography' – that will help contextualize our results and better understand their implications for policy. Let us begin by thinking of agglomeration as a function of increasing returns to scale (IRS), product market competition, factor market competition (especially land and labor) and trade/transportation costs (for example, Krugman & Venables 1990). Two main types of forces are in play: centripetal forces, which bring economic activities together, and centrifugal forces, which push them apart. The equilibrium level of agglomeration is the result of a balance between these two forces. If trade costs are not negligible, the presence of IRS will tend to pull economic activities towards agglomeration. To see this, let us suppose that there are two regions: one large or rich region - let us call it "the core" - and another small/poor region - "the periphery". Firms will have a tendency to locate in the core region where they can take advantage of larger market access as they can reach more customers with lower trade costs. This results in greater efficiency and profits due to IRS and thus greater entry in the core (that is, agglomeration) - as compared with firms producing in the smaller market. At the same time, competition among firms on product and factor markets tends to counteract those agglomeration forces. Indeed, firms will be pressured to move out of "the core" to alleviate the effects of tougher competition (lower product prices), and tougher factor markets (higher factor prices), both resulting in lower profitability. Since at least Krugman and Venables (1990), the effect of trade/transportation costs on the degree of agglomeration has been known to be ambiguous as theoretical models of agglomeration can yield multiple equilibria. Therefore, we will devote the rest of this section to a discussion of the circumstances under which declining trade costs result in higher or lower levels of agglomeration.

To this end, we draw a distinction that is inspired by the theoretical model of Puga (1999). Consider two alternative cases, one in which labor is mobile across regions, and another in which labor is immobile. If one area (the core) initially has a greater population, firms there will have higher productivity and higher real wages (due to IRS), which will lead workers to migrate in, leading to an even greater population and yet more productive firms, while diminishing wage differentials. Thus, interregional labor mobility is likely to fuel agglomeration. If labor is immobile, so that workers cannot

migrate in response to higher real wages in the core, the tendency towards agglomeration is choked off, as the denser region does not absorb more resources and firms must pay higher real wages for labor, which in turn discourages the entry of new firms.

Labor mobility should therefore be associated with higher levels of agglomeration, while labor immobility should lead to lower levels of agglomeration. Moreover, Puga (1999) also shows that if labor is particularly immobile, so that real wages are very high in the core, lowering trade costs may lead to even less agglomeration. This is because lower trade costs allow firms to move away from the core without sacrificing their market share – it is more profitable to (cheaply) transport your products to the larger market while taking advantage of the lower factor costs in periphery regions. If labor is very mobile, one may see the opposite relationship: declining trade costs may lead to greater agglomeration, as labor mobility prevents factor prices from increasing in the core so that firms can settle there without negative consequences.

The preceding discussion followed Puga (1999) in focusing on a distinction between labor mobility versus labor immobility, but we can broaden the concept to include other factors relevant to the Indian context. In particular, consider forces such as "congestion effects" due to poor within city transport infrastructure or poor land management policies. These forces also prevent labor and land prices from being equalized across core and periphery regions, and, hence, will have the same effects on agglomeration as labor immobility. The forces discussed thus far are not introduced arbitrarily: labor immobility, poor within city transport infrastructure and poor land management policies are considered by many to characterize the Indian economy in its recent past and present (see, for example, Ahluwalia et al. 2014; McKinsey Global Institute 2010). If this is so, we should expect India to have relatively low levels of agglomeration, and we should further expect reductions in interregional trade costs to lead to yet lower levels of agglomeration.

6 Results

6.1 Industry Level Agglomeration Indices

Table 2 presents summary statistics for both weighted and unweighted EG indices over time. The data show that there has been a significant decrease in average spatial concentration over time. 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 (SWEGI), we see that the decline is profound, from a high of 0.114 in 1998 to a low of 0.028 in 2013. There has been a corresponding reduction in 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 0.05 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]

[Table 2 around here]

However, these economy-wide averages mask significant sectoral differences. Figure 1 plots the median values of the SWEGI separately by major sector, distinguishing industries in the secondary sector from those in the tertiary sector.[11] Several facts are apparent from this figure. 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 general decline in average levels of concentration, the biggest portion of the economy-wide decline in concentration is accounted for by the secondary sector. Within the secondary sector, it turns out that the results vary markedly by capital intensity of the industry: Figure 2 shows that the decline in concentration was driven by a precipitous drop in concentration among capital intensive firms in particular.[12]

[Figures 1 and 2 around here]

[10] As we noted earlier, one potential weakness of the EC is that it may undercount the very smallest informal establishments. To ensure that changes in the degree of undercounting over time are not driving our intertemporal results, we include the following robustness check: we regenerate the EGI omitting all establishments with less than 5 workers, since it is establishments of this size that are potentially undercounted. The results, displayed in Table S1 in the online supplementary material, show that the intertemporal patterns observed in our main results are not sensitive to the inclusion or exclusion of the very smallest establishments.

[11] An analogous figure for the unweighted EGI is provided in Figure S1 in the online supplementary material.

[12] We use data from the 1994/95 and 2009/10 ASI and NSSO surveys to classify industries according to their average capital/labor ratios.

Further insights into the patterns of the spatial distribution of industries can be obtained from Tables 3 and 4, which list the top 10 most and least concentrated industries as measured by the SWEGI in the years 1998 and 2013 (that is, the endpoints of our data). What is immediately clear from Table 3 is that manufacturing industries dominate the list of most concentrated industries, a fact consistent with Figure 1. We also observe significant persistence in which industries are most concentrated over time, with, for example, electrical machinery, rubber and plastic, motor vehicles, and radio, television and communication among the most concentrated industries throughout the period. As this list suggests, most of the heavily concentrated industries are high value-added industries with the technology of the 'second industrial revolution' and ICT era.[13] The least spatially concentrated industries are mostly services (see Table 4), in particular those related to retailing (for example, sales of motor vehicles), and labor intensive services with high transport costs such as hotels and restaurants, education, and repairs.

[Tables 3 and 4 around here]

6.2 Coagglomeration

Ellison and Glaeser's (1997) coagglomeration index offers a way to measure the extent to which plants in one industry locate in the same regions as plants of a different industry. Table 5, which lists the 10 industry pairs with the highest coagglomeration indices, shows that high-tech, ICT-based manufacturing and service industries demonstrate the highest tendency for coagglomeration, especially in 2013. The strength of this pattern is highly suggestive of knowledge or technology spillovers across even relatively broadly defined 2 digit industries and which have grown in importance over time.

[Table 5 around here]

[13] Interestingly, the most concentrated industries among services are also ICT related industries such as computer related industries, and post and telecommunication. In contrast, the least concentrated manufacturing industries are those which can be characterized as traditional industries – or industries of the first industrial revolution – such as textiles, apparel, food products and metal products. Tables S2-S5 in the online supplementary material show the most and least concentrated industries within manufacturing and service industries separately.

7 Policy Implications and Discussion

In this concluding section, we devote some space to a discussion of the possible causes and potential implications of what is perhaps the paper's most striking finding: that industrial agglomeration in India has decreased substantially between 1998 and 2013. This decline in industrial agglomeration is important from a policy perspective, as a good deal of research demonstrates that agglomeration plays an important role in boosting firm productivity and economic growth due to various positive externalities such as knowledge spillovers, labor market pooling and input-output linkages across firms in industrial clusters (for example, Duranton 2015). If the level of agglomeration in India is suboptimal, it could mean that India is losing out an important source of economic growth.

Two findings from the literature provide a prima facie reason for believing that this may be the case. First, and with the caveat that the EGI may not be reliable when compared across countries using different spatial units, the literature documents substantially higher EGIs in developed countries than those we have estimated for India. Second, data from another large developing country, China, suggest that it has experienced increasing industrial concentration while growing even faster than India (Lu & Tao 2009). These findings are not conclusive, but they do suggest that agglomeration may be moving in the wrong direction in India.

What might explain the decline in industrial concentration? We believe that at least two sets of factors may be relevant: 1) specific industrial policies; and 2) a combination of decreasing transportation costs in the presence of significant frictions in factor markets. First, notwithstanding the industrial delicensing reforms of 1991 – whereby proposals for industrial investment in so-called backward areas were encouraged while those located in or around metropolitan areas were discouraged (Fernandes & Sharma 2012) – industrial policy in India has continued to provide explicit incentives for the dispersal of industry. For example, India's central government initiated a program in 1994 that identified 123 industrially backward districts out of 360 districts belonging to 14 major states of India and offered tax exemptions to new industrial firms located in those districts. The results of Hasan et al. (2017), who evaluate the program using a regression discontinuity design, are consistent with the idea that it

contributed to dispersal of manufacturing, with the relatively more advanced among the backward districts experiencing large increases in numbers of firms and employment by 1998. Similarly, the central government implemented a tax incentive and capital subsidy scheme in 2003 in two relatively under-industrialised states (Himachal Pradesh and Uttarakhand). Applying a difference-in-differences approach, Chaurey (2016) finds that the policy resulted in large increases in outcomes such as employment, number of firms, and total output in the treatment states relative to the control states.

Our second explanation for the declining geographic concentration relies on the interpretative framework discussed in Section 5. In that framework, we argue that if there are significant frictions in the factor markets for land and labor, reductions in interregional transportation costs are likely to lead to greater dispersion of economic activity, as the lower trade costs allow firms to alleviate factor market competition by moving away from congested areas.

How relevant are these conditions? Beginning with labor mobility, there is an extensive literature which documents that internal migration is low in India. Based on Bell et al.'s (2015) estimates of internal migration rates in 80 countries over a five-year interval between 2000 and 2010, India is found to have the lowest migration rate. Similarly, Kone et al. (2017) note that although internal migrants represented 30 percent of India's population in 2001, two-thirds were migrants within districts, and more than half were women migrating for marriage. They also note that in comparison to India, internal migration rates across states were nearly four times higher in Brazil and China, and more than nine times higher in the United States in the five years ending in 2001, despite the fact that in countries including China, urban-rural wage gaps are considerably lower than in India.

Turning to the issue of land, perhaps the key non-tradeable input for any kind of economic activity, there are good reasons to believe that India's policy and regulatory frameworks have hindered land markets and obstructed access to the large tracts of land essential for industrial agglomerations to form. Regulatory barriers in rezoning land use from agriculture to non-agricultural activities have made it difficult for the development of industrial clusters in rural areas, while land ceiling regulations, restrictive building codes, and very low floor area ratios, have arguably made urban land even more scarce than it need be (Sridhar 2010; Brueckner & Sridhar 2012).

Last, the period of our study (1998 – 2013) coincided with significant improvements in transport connectivity between cities and in rural areas thanks to large investments in three national highway systems (the Golden Quadrilateral (GQ), the North-South highway, and the East-West highway) and in rural roads (PMGSY). Thus, the conditions outlined in our interpretive framework seem to have been satisfied over this period: reductions in trade costs alongside limited labor mobility and policy-induced scarcity of land should, in theory, have led to the dispersion of industry.

Is this what really happened? It is beyond the scope of this paper to provide a definitive answer but we can nevertheless provide some suggestive evidence that is consistent with the hypothesis. To do so, we divide up the states of India first according to whether they had efficient or inefficient land management policies, and then according to whether they benefited from having a national highway built through their territory or not. For each partition, we separately compute and compare EGIs for the industries located in those partitions. Our partition of states by land management regime is based on Hasan et al. (2018).[14]

Strictly speaking, the fact that Indian firms may be able to locate across either of 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 - presented in Tables 6 and 7 - are suggestive. We find that states with inefficient land management policies started off in 1998 with much higher concentration than in states with efficient policy regimes. Over time, however, concentration in those states fell - so much so that average concentration is now lower in inefficient land management states than in efficient ones (Table 6). The results for transportation are similar: states that became directly connected to the

[14] The partition divides states according to their success on seven different measures that are relevant to urban land management. These seven 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 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.

Golden Quadrilateral (GQ) highway network between 1998 and 2013 saw their industrial concentration decline much faster than states that were not intersected by the GQ (Table 7).[15]

[Tables 6 and 7 around here]

These findings are consistent with the hypothesis that inefficient land management policies and improved transport infrastructure may be driving the trend towards de-agglomeration – although we realize that the partitions generating the results are endogenous to the process of economic development and are thus not conclusive. If correct, the hypothesis implies that forces of congestion are preventing a more efficient level of agglomeration from taking place, and there may thus be a role for policy to reverse this trend by reducing inefficiencies in land markets and encouraging greater labor mobility.

We conclude with an observation about the likely impact of one of the present government's signature policies: the Goods and Services Tax (GST). To the extent that this – or any other future reforms – help create a common market by reducing trade costs – but without addressing inefficiencies in the markets for land and labor – we should expect yet greater dispersion of economic activity, even though it may involve firms missing out on locating in areas where agglomeration economies could be reaped. 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, and more conclusively test our hypothesis that congestion costs related to land and labor market frictions are disrupting the forces of agglomeration, at the potential cost of future economic growth.

[15] The results are similar when partitioning states according to other highway networks built over the period, or when using the unweighted EGI.

References

Ahluwalia I.J., Kanbur R. & Mohanty P.K. (eds) (2014). Urbanisation in India: Challenges, Opportunities and the Way Forward. New Delhi: SAGE Publications.

Amirapu A. & Gechter M. (Forthcoming). Labor regulations and the cost of corruption: Evidence from the Indian firm size distribution. Review of Economics and Statistics.

Bell M., Charles-Edwards E., Ueffing P., Stillwell J., Kupiszewski M. & Kupiszewska D. (2015). Internal migration and development: Comparing migration intensities around the world. Population and Development Review, **41** (1), 33-58.

Brueckner J.K. & 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.

Chaurey R. (2016). Location-based tax incentives: Evidence from India. Journal of Public Economics, **156,** 101-120.

Ciccone A. & Hall R. (1996). Productivity and density of economic activity. American Economic Review, **86** (1), 54-70**.**

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

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

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.

Duranton G. (2015). Growing through cities in developing countries. The World Bank Research Observer, **30** (1), 39–73.

Ellison G. & Glaeser E.L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy, **105** (5), 889-927.

Fernandes A.M. & Sharma G. (2012). Together we stand? agglomeration in Indian manufacturing. World Bank Policy Research Working Paper no. 6062.

Ghani E., Goswami A.G. & Kerr W. (2016). Spatial development and agglomeration economies in services: Lessons from India. World Bank Policy Research Working Paper no. 7741.

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

Hasan R., Jiang Y. & Kundu D. (2018). The growth of Indian cities and "Good" Jobs: Evidence from the 2000s. forthcoming India Policy Forum. NCAER. Sage Publications.

Hasan R., Jiang Y. & Rafols R. (2017). Place-based preferential tax policy and its spatial effects: Evidence from India's program on industrially backward districts. ADB Economics Working Paper no. 524, Asian Development Bank (ADB).

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

Kone Z.L., Mattoo A., Ozden C. & Sharma S. (2017). Internal borders and migration in India. World Bank Policy Research Working Paper no. 8244.

Krugman, P. (1991). Geography and Trade. Cambridge, Massachusetts: The MIT Press.

Krugman P. & Venables A.J. (1990). Integration and the competitiveness of peripheral industry. In: De Macedo JB & Bliss C (eds), Unity with Diversity in the European Economy: The Community's Southern Frontier. (pp. 56-75). Cambridge: Cambridge University Press.

Lall S.V., Shalizi Z. & Deichmann U. (2005). Agglomeration economies and productivity in Indian industry. Journal of Development Economics, **73** (2), 643-673

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

Manna G.C. (2010). Current status of industrial statistics in India: Strengths and weaknesses. Economic and Political Weekly, **XLV** (46), 67-76.

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

McKinsey Global Institute. (2010). *India's* Urban Awakening: Building Inclusive Cities Sustaining Economic Growth. New York: McKinsey & Company.

Panagariya A. (2008). India: The Emerging Giant. New York, New York: Oxford University Press.

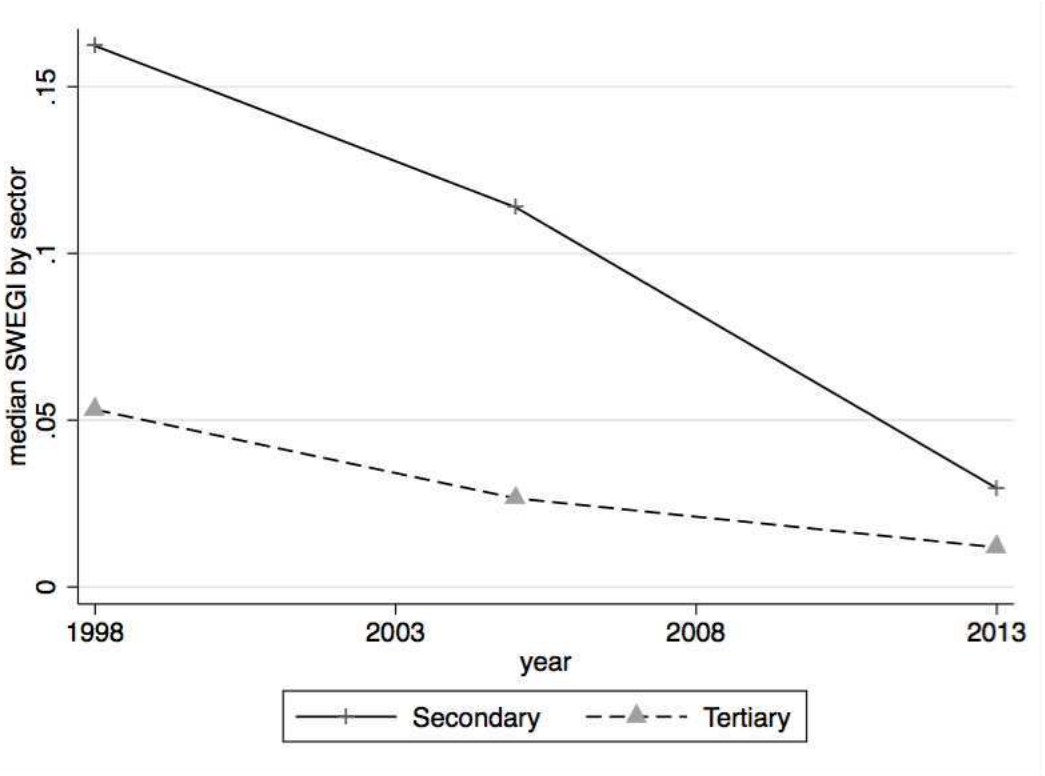
Puga D. (1999). The rise and fall of regional inequalities. European Economic Review, **43** (2), 303-334.

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

Unni J. & Raveendran G., (2006). Are the results of the economic census robust? Economic and Political Weekly, **XLI** (33), 3558-3560.

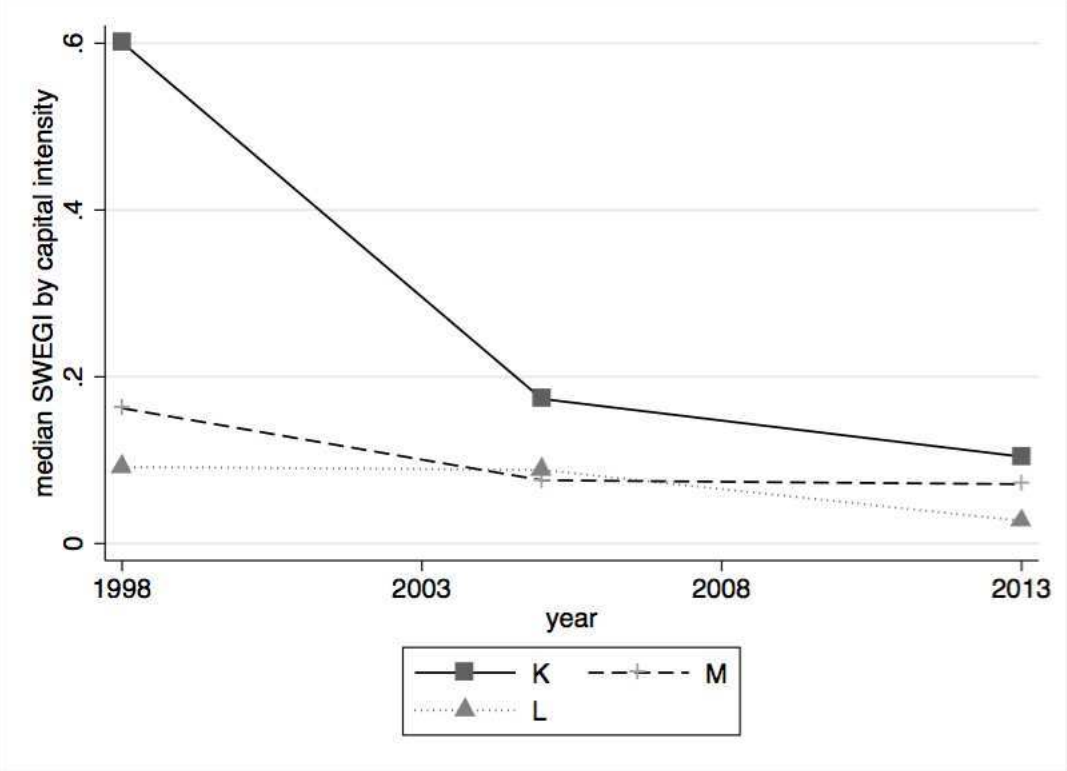
Figures and Tables for "Geographic Concentration in Indian Manufacturing and Service Industries: Evidence from 1998 - 2013"

Figure 1: Median Spatially Weighted EGIs Over Time by Sector



Note: This figure displays median values of a spatially-weighted Ellison Glaeser Index (SWEGI) of concentration for 2 digit industries in the secondary and tertiary sectors. Source: 1998, 2005 and 2013 Economic Censuses of India.

Figure 2: Median Spatially Weighted EGIs Over Time by Capital Intensity of Industry



Note: This figure displays median values of a spatially-weighted Ellison Glaeser Index (SWEGI) of concentration for 2 digit industries in the secondary sector. Industries are grouped according to whether they are capital intensive (K), labor intensive (L), or neither (M). These divisions are based on industries' average capital-labor ratios using data from the 1994/95 and 2010/11 Annual Survey of Industries (ASI) and NSS Unorganized Sector Surveys. Source: 1998, 2005 and 2013 Economic Censuses of India.

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 Agglomeration Indices Over Time

Measure of Agglomeration	Statistics	1998	2005	2013
Ellison and Glaeser Index (EGI)	median	0.015	0.009	0.007
	mean	0.035	0.017	0.018
	std dev	0.063	0.018	0.029
	share $>$ 0.05	0.186	0.102	0.089
	obs	59	59	56
Spatially Weighted EGI (SWEGI)	median	0.114	0.061	0.028
	mean	0.386	0.111	0.088
	std dev	0.783	0.131	0.170
	share $>$ 0.05	0.695	0.576	0.339
	obs	59	59	56

Table 3: 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 4: 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 5: 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 6: Average SWEGIs in Efficient vs Inefficient Land Management Regions

Year	*1998*	*2005*	*2013*
Efficient Land	0..070	0.053	0.065
Management	(0.195)	(0.072)	(0.102)
Inefficient Land	0.111	0.058	0.051
Management	(0.174)	(0.073)	(0.063)

Note: This table displays the mean and standard deviation over time of a spatially-weighted Ellison Glaeser Index (SWEGI) of concentration among 2 digit industries, calculated separately for states in India with "efficient" vs "inefficient" land management regimes. More information about the categorization, which is based on Hasan, Jiang, and Kundu (forthcoming), can be found in the text. Source: 1998, 2005 and 2013 Economic Censuses of India.

Table 7: Average SWEGIs by Golden Quadrilateral Status

Year	*1998*	*2005*	*2013*
Connected to	0.556	0.172	0.133
GQ	(1.016)	(0.206)	(0.254)
Not Connected	0.084	0.067	0.058
to GQ	(0.168)	(0.107)	(0.093)

Note: This table displays the mean and standard deviation over time of a spatially-weighted Ellison Glaeser Index (SWEGI) of concentration among 2 digit industries, calculated separately for states in India that were intersected by some part of the Golden Quadrilateral (GW) national highway network. Source: 1998, 2005 and 2013 Economic Censuses of India.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website

Figure S1 Median EGIs over time by sector

Figure S2 Median EGIs over time by capital intensity of industry

Table S1 Summary statistics for agglomeration indices over time (excluding establishments with < workers)

Table S2 Highest gammas (most localized) in manufacturing

Table S3 Highest gammas (most localized) in services

Table S4 Lowest gammas (least localized) in manufacturing

Table S5 Lowest gammas (least localized) in services