Environment and Planning A 2005, volume 37, pages 331 – 353

DOI:10.1068/a373

Do localization economies matter in cluster formation? Questioning the conventional wisdom with data from Indian metropolises

Sanjoy Chakravorty
Department of Geography and Urban Studies, 310 Gladfelter Hall, Temple University, Philadelphia, PA 19122, USA; e-mail: sanjoy@temple.edu

Jun Koo
Cleveland State University, Maxine Goodman Levin College of Urban Affairs, Cleveland, OH 44115, USA; e-mail: jkoo@urban.csuohio.edu

Somik V Lall
Development Research Group, The World Bank, Washington, DC 20433, USA; e-mail: slall1@worldbank.org
Received 7 January 2004; in revised form 30 March 2004

Abstract. The large and growing literature on industrial clustering suggests that firms seek locations that provide localization economies (benefits from having common buyers and suppliers, a specialized or skilled labor pool, and informal knowledge transfers). This study of manufacturing industry clusters in three Indian metropolises suggests instead that industry location decisions are guided by market imperfections, specifically rigidities in the land market caused by state action (segregationist or environmental policies, the absence of exit policies, and activist industrial promotion policies). For the investigation the authors use geographically disaggregated industry location and size data from Mumbai, Kolkata, and Chennai, to analyze eight industrial sectors (food and beverages, textiles, leather, printing and publishing, chemicals, metals, machinery, and electrical and electronics). The authors test for evidence of global and local clustering and distinguish between and test for coclustering and colocation of industries. The results are indicative rather than absolute and suggest that for location decisions general urbanization economies are more important than are localization economies.

Industrial clusters have reemerged as important objects of research and policy analysis. The benefits of industry clustering were identified early on by Alfred Marshall (1919) who suggested that these arise from localization economies: namely, the availability of common buyers and suppliers, the formation of a specialized or skilled labor pool, and the informal transfer of knowledge (on trade secrets, production processes, market agents etc). Krugman's (1991; 1996) work in economic geography and Porter's (1990; 1996) work in business economics have drawn the interest of economists to the idea of 'increasing returns' to proximity in the form of clusters (see Fujita et al, 1999). Meanwhile, a tradition of studying the locational aspects of economic activity exists in several academic disciplines (see Greenhut and Greenhut, 1975; Hotelling, 1929; Isard 1956; Lösch, 1956; Weber, 1929; for a survey, see Walker, 2000). The literature on industrial clustering and its causes (localization and/or urbanization economies, proximity to other firms and/or consumers) and effects (economic growth, unbalanced development, regional inequality, and global industrial restructuring) continues to proliferate in journals of geography, economics, planning, and development (for empirical work in developing nations, see Chakravorty, 2000; Henderson and Kuncuro, 1996; Lee, 1989; Nadvi and Schmitz, 1999; World Bank, 1999).

However, some fundamental issues remain unresolved. In this paper we address one of these issues in one country—namely, we address the relative importance of localization and urbanization economies in cluster formation in Indian metropolises. The analytical problems begin with the fact that there is little agreement on what a cluster is in empirical terms: how many firms constitute a cluster? How many workers should a cluster have? What is the geographic area over which we count these numbers? How close is close enough, how far is too far? According to Martin and Sunley (2003, page 12) "to use the term [cluster] to refer to any spatial scale is stretching the concept to the limits of credulity, and assumes that 'clustering processes' are scale-independent." It seems to us that the constituent elements of agglomeration economies—which are localization and urbanization economies—conceptually account for most scale issues. Localization economies refer to features that are very proximate, being in the neighborhood, whereas urbanization economies refer to features that are proximate, but less so, being in the region rather than in the neighborhood. This is far from precise, because neighborhood and region remain undefined. Later we will be more specific.

The analytical problems continue with the implicit assumption in the clustering literature that industry location decisions are made in flexible, if not unfettered, land markets; that decisionmakers at the firm level choose locations from a multitude of options. This is a critical assumption because a firm's location decision is fundamentally a choice in the land market. We know that even in strongly market-oriented economies (such as the United States) there are land-market constraints such as zoning. In fact, clustering policies à la Porter rely on being able to use state power to intervene in land markets. We argue that land-market rigidities exist everywhere; in the specific case of Indian metropolises, these rigidities seriously constrain location choices for firms, so much so that localization economies practically do not matter when the location choice is made. The rigidities in the land market arise from state actions, specifically industrial location policy, environmental policy on segregating or buffering out polluting industry, and land-use policy on land-use change. This does not invalidate the possibility of realizing localization economies after the location decision; in fact, it is quite possible that the land-market rigidities aid clustering, which may lead to localization economies during the production phase. However, the evidence indicates that most manufacturing firms have relatively few location choices when they enter the market, and probably the only external economies they can realize with certainty are general urbanization economies. At the point of the location decision, the region matters more than the neighborhood.

We reached this conclusion after systematically analyzing disaggregated manufacturing industry location data for eight industry sectors in three Indian metropolises: Mumbai, Kolkata, and Chennai (these are the new names of Bombay, Calcutta, and Madras, respectively). As data at this level of geographic detail have never been analyzed in India, we begin by asking the basic empirical questions for specific industries: Where do industries locate within a metropolitan area? Do industries cluster? Do different industrial sectors have different patterns of location or clustering? Can these patterns be understood with reference to industry or firm characteristics? We then consider questions relating to the spatial relationships between the different industry sectors; specifically, we ask: Do industries colocate or cocluster? Through all this, our goal is to answer questions on location theory: What is the role of localization economies in cluster formation in metropolises? To what extent do these derive from interfirm transactions such as collaboration and information sharing, or intraindustry transactions such as through shared buyer–supplier networks, intermediate goods, and specialized labor pools? What is the role of urbanization economies as manifested in regional industrial diversity, general (as opposed to specialized) labor pools, and buyers and suppliers and knowledge transfers in the region as opposed to the neighborhood?

In the following pages we first provide the necessary background: the theory of industrial clustering, the data used for the analysis, and cluster measurement methods.

Next, we test eight industrial sectors (food and beverages; textiles; leather; printing and publishing; chemicals; metals; machinery; electrical and electronics) for evidence of global and local clustering (explained later), and distinguish between and test for coclustering and colocation of industries (also explained later). The results suggest a temporal model of industry location in mixed rather than specialized industrial districts. There is little evidence that localization economies from labor markets or buyer – supplier networks drive industrial location decisions. In the concluding section we detail the argument that land-use policies—segregationist or environmental policies, the absence of exit and land-use change policies, and activist industrial promotion policies—are the key influences on the intrametropolitan spatial distribution of industry.

Background

Why do industries cluster?

Clustering is a term describing a phenomenon in which events or artifacts are not randomly distributed over space, but tend to be organized into proximate groups. Industrial clustering is a process that has been observed from the beginning of industrialization. From the cotton mills of Lancashire and automobile manufacturing in Detroit, to the textile mills of Ahmadabad and Mumbai and the tanneries of Kolkata and Arcot, even the casual observer can visually identify industry clusters. It seems obvious that competing firms in the same industry derive some benefit from locating in proximity to each other. The benefits that are external to the firm and accrue to similar firms in proximity are called the *economies of localization*. Now, these typically are not the only firms in the immediate region. There are usually other factories, producing other and similar goods, distributed through other and similar channels, for other and similar markets. These other firms, and their employees, and the service workers who provide food, education, transportation, and healthcare for all these employees and their families, form, typically, an urban area. All the firms that benefit from being in the urban area, regardless of whether or not there are other similar firms in the area, derive *economies of urbanization* from their location choice.

To put it in another way, at the *firm level*, it is expected that the size and number of firms (that is, the competitive structure) will influence internal returns to scale. In particular, as demand for a firm's goods and services increases (say, as a result of improved access to consumer markets), the entrepreneur has an incentive to increase the scale of production by restructuring the production process through the use of specialized workers and investing in cost-reducing technologies (Lall et al, forthcoming). At the *industry level* we expect to see quantifiable localized benefits of clustering that accrue to all firms in a given industry or in a set of interrelated industries. Productivity is likely to be higher in regions where an industry is more spatially concentrated because of the increased potential of knowledge spillovers and dense buyer – supplier networks, access to a specialized labor pool, and opportunities for efficient subcontracting. Last, at the *metropolitan area level*, economies of scale result not from the size of a specific industry or market but from the overall size, diversity, and spatial configuration of the metropolitan area. These economies of urbanization include access to specialized financial and professional services, availability of a large labor pool with multiple specializations, interindustry information transfers, and the availability of less costly general infrastructure (see Parr, 2002). At the interregional scale these gains are expected to lead to industry concentration in metropolitan and other leading urban regions (as a result of urbanization economies); at the metropolitan scale the gains from localization economies are expected to lead to the creation of local industrial clusters.

These typically unquantified agglomeration economies are one set of inputs into the location decision of a firm. There are other significant factors that a firm facing a location decision must consider. The two most important of these additional factors (especially in developing countries) are the availability of infrastructure and the regulatory framework, both arenas where the state is the key player. The state not only sets the rules of market entry and participation but also is the primary, often the sole, provider of physical and social infrastructure and is often directly active in the production process. In India, in its efforts to capture the 'commanding heights of the economy', the state invested heavily in capital-intensive industry, such as integrated steel and power plants. It was so successful in its efforts that as recently as the late 1990s nine of the top ten, and twenty of the top twenty-five, corporations in India were public sector units (Nayar, 1998). At the local level, the state's regulatory role goes beyond setting the rules of market participation; by being the single largest owner of land, by having the police and taking powers to acquire necessary land, and by being the final arbitrator on land-use decisions, the state, as we shall show later, has a very strong influence on industrial location decisions within metropolitan areas.

Data issues

In order to undertake submetropolitan-level analysis it is necessary to have spatially disaggregated data. In India, industrial data are collected by the Central Statistical Organization (CSO) and disseminated as the Annual Survey of Industries (ASI). In the late 1990s the ASI data were first released at the district level and then at the firm level. These data are collected from a survey undertaken by CSO on a sample taken from an industry sampling frame that includes every registered (or legal) industrial unit with at least ten workers. This sampling frame contains one record for each industrial unit and includes three critical pieces of information: the National Industrial Classification (NIC) code, the number of workers in the unit, and the street address, with, sometimes, a pin code (equivalent to US zip codes).

The last piece of information is the key to disaggregating the district data down to smaller enumeration units. We have access to the sampling frame for the whole country for the enumeration period 1998–99. On further examination we found that, although street addresses were generally available for all metropolitan areas, there were no base maps of streets to which these addresses could be matched. Hence we had to rely on the pin code information, which, however, turned out to be erratically available. For some cities the pin codes were generally available or imputable, for other cities they simply were not available. We identified Mumbai, Kolkata, and Chennai as the three metropolises with enough information to begin geocoding the industry location data to the pin code level.

The pin code maps were acquired from a private sector firm in New Delhi (ML Infomap, 1998). These maps have somewhat variable coverage. For Kolkata and Chennai these maps cover the largest definition of their metropolitan areas. For Mumbai the pin codes cover the district of Greater Bombay only; that is, the far northern and eastern suburban reaches of the Mumbai metropolitan area (in Thane and Raigad districts) are not covered. Even in the covered area there appear to be some situations where adjoining pin codes have been merged. As a result, some data that are known to be in the metropolitan area could not be geocoded to pin codes. We have been able to achieve the following 'hit' rates, that is successfully geocoded factory records: in Mumbai 99.9%, in Kolkata 97.1%, in Chennai 97.5%. According to Ratcliffe (2004), a 'hit rate' of 85% is acceptable for most map-based analysis; thus we believe that these are acceptable levels of address matching.

Finally, we aggregated the firms into eight distinct and internally consistent sectors, as follows [the NIC codes listed in parenthesis generally correspond to the international standard industrial codes (ISICs) or 1987 Standard Industrial Codes (SICs) used in the USA]:

1. food processing (151, 152, 153, 154, 155);
2. textiles and textile products, including wearing apparel (171, 172, 173, 181);
3. leather and leather products (191, 192);
4. paper products, printing and publishing (210, 221, 222);
5. chemical, chemical products, rubber and plastic products (241, 242, 243, 251, 252);
6. basic metals and metal products (271, 272, 273, 281, 289);
7. mechanical machinery and equipment (291, 292);
8. electrical and electronics (including computer) equipment (292, 300, 31, 32).

The measurement of clustering

Spatial statistics are the most widely used tools for identifying and analyzing spatial patterns (for excellent discussions on the subject, see Anselin, 1995; Getis and Ord, 1992). In classifying spatial patterns, researchers are often interested in determining whether the distribution of activity is clustered, dispersed, or random. At this point, it is useful to differentiate between industrial 'clustering' and 'concentration'. Several devices to measure industry concentration have recently been operationalized. The 'spatial Gini' and γ (from Ellison and Glaeser, 1997) have become quite well known. These measures, and older ones such as the location quotient (discussed later), suffer from a common drawback, one that White (1983) termed the 'checkerboard problem', whereby these measures are not really spatial—any geographical arrangement of parcels (in this case pin codes) would yield the same measure of concentration. Hence 'concentration' has to be distinguished from 'clustering', where clustering is explicitly spatial; that is, geographical arrangements are incorporated in measures of clustering, but not in measures of concentration.

Clustering is best understood in the context of spatial autocorrelation, a term that describes conditions where the attribute values being studied are correlated according to the geographic ordering of the objects. When the location of firms is spatially autocorrelated, it implies that the geographic distribution of economic activity is not random and is likely to be determined by some underlying political, economic, or physical factors attributable to each geographical unit. Hence, strong positive spatial autocorrelations mean that the attribute values of adjacent geographical units are closely related.

One of the most popular measures of spatial autocorrelation is Moran's *I*. There are two types of Moran's *I*. The global Moran is a measure describing the overall spatial relationship across all geographical units. Therefore, only one value is derived for the entire study area. The local Moran (often called the local indicator of spatial association, or LISA) is a measure that is designed to describe the heterogeneity of spatial association across different geographical units. The global Moran can be thought of as a 'regional' measure, the local Moran as a 'neighborhood' measure.

The global Moran's I can be defined as

$$I = \frac{n \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W \sum_i (x_i - \bar{x})^2}, \tag{1}$$

where n is the number of areal units, x_i is the value of the variable of interest in areal unit i where there are j areal units, w_{ij} is a weight derived from a spatial weight matrix, and W is the sum of all cell values of the weight matrix. For the calculation of Moran's *I*, both a binary and a stochastic weight matrix can be used. A binary weight matrix defines the connectivity of pairs of regions with 0s and 1s. When two regions are adjacent, the corresponding cell value is 1, but otherwise it is 0. In contrast, the stochastic weight matrix takes into account the number of immediate neighbors. Rather than assigning 1 to every neighbor, '1/(total number of neighbors)' is used as a weight for areal unit *i*.

Equation (1) shows that the calculation of the global Moran's I is based on a comparison of values in neighboring geographical units. If neighboring units have similar values over the entire study area, the statistic will show a strong positive spatial association. If dissimilar values are observed among neighboring units, the statistic should indicate a strong negative spatial association. However, the magnitude of spatial association is not necessarily or usually uniform over the space. It is more likely to be heterogeneous according to local characteristics that influence the formation of spatial structure. The local Moran's I can be used as an indicator of heterogeneity in spatial association over geographical units and is defined as

$$I_i = z_i \sum_i w_{ij} z_j \,, \qquad z_i = \frac{(x_i - \bar{x})}{\delta} \,, \tag{2}$$

where z_i is the deviation from the mean, and δ is the standard deviation of x_i. Similar to the case for the global Moran, a high value of local Moran implies the association of similar values, whereas a low value means the association of dissimilar values. It is important to note that local instabilities or clusters may exist in a given spatial distribution even when the distribution as a whole does not exhibit a statistically significant level of spatial autocorrelation.

This analytical method automatically answers a key question identified at the beginning of this paper: what is a neighborhood, or over what geographical area do we calculate clusters? In most cases, a neighborhood is defined as the sum of a subregional areal unit and its adjacent (or first-order) neighbors. In this case, it is a pin code and its contiguous pin codes. It is possible to use other formulations of neighborhood. We can use distance-based measures, whereby every areal unit for which the centroid falls within a chosen distance of the centroid of a given areal unit becomes the neighbor of that unit. Also, we can use not only immediately adjacent neighbors but also second-order neighbors, that is, the immediate neighbors of immediate neighbors. Use of this definition would form larger neighborhoods. There is no theoretical guidance on which of these measures is appropriate for which kind of analysis. We used first-order neighbors and a distance cut-off definition of neighborhoods. Only the contiguity-based (or first-order) local Morans are reported here. The distance-based measure yields similar results.

Analysis

We have analyzed data for eight industrial sectors in three metropolitan areas for two variables (number of factories and number of workers). This means that for each basic test forty-eight result points are presented. There is some danger of 'drowning in data' because the results vary by industry sector, by metropolis, and for factories and workers. We attempt to simplify some of the 'data clutter' in the discussion accompanying the tables and maps. It is important to note that from a theoretical perspective the most significant variable is the factories–workers distinction because it is based on the assumption that factory size has a significant bearing on location decisions. Large factories can be expected to rely more on internal economies of scale; small factories may rely on external or localization economies. Here the variable 'factories' suggests small-scale units; when factories are clustered it suggests that a large number of small firms are clustered. Conversely, when workers are clustered we can assume that a small number of large firms are clustered.

We present the results of the analysis in three parts. First, we test for global clustering; that is, we examine whether the distribution of factories and workers in the metropolis as a whole is clustered. Second, we test for local clustering by using maps identifying local clusters and a table summarizing the map information. Third, in the longest section, we test for colocation and coclustering of industry pairs. Each section contains findings that inform both the empirical and the theoretical questions identified in the introductory section.

Global clustering

The measures of global clustering (Moran's I) for the eight industry sectors for the three metropolises are reported in table 1. In general, clustering is most consistent in Kolkata; clustering in Chennai is more evident than in Mumbai, with more consistent clustering among factories than workers in Chennai. Clustering among factories and workers in the same industry and city is a common pattern; nine of the twenty-four pairs show this combination. However, combinations with only one of them (factories or workers) being clustered is equally commonly observed; in seven cases only factories are clustered, and in two cases only workers are clustered. This suggests, as expected, that small units are more clustered than are large units. However, at this point in the analysis, the overall picture is unclear.

Table 1. Indices of global clustering (data source: CSO, 1998–99).

	Factories[a]			Workers[a]		
	total number	I	Z of I	total number	I	Z of I
Mumbai						
Food and beverages	274	0.172*	2.951*	16 992	0.156*	2.699*
Textiles	1 988	0.100	1.797	131 318	0.191*	3.259*
Leather	115	0.074	1.374	2 629	0.018	0.474
Printing and publishing	882	0.060	1.139	17 001	0.048	0.959
Chemicals	1 146	0.149*	2.591*	41 470	0.008	0.031
Metals	894	0.069	1.290	25 056	−0.019	0.145
Machinery	767	0.104	1.867	40 808	−0.045	0.560
Electrical and electronic	802	0.244*	4.133*	58 871	0.141*	2.455*
Kolkata						
Food and beverages	355	0.165*	3.332*	9 573	0.170*	3.429*
Textiles	355	0.063	1.354	90 621	0.168*	3.380*
Leather	283	0.211*	4.216*	6 149	0.316*	6.229*
Printing and publishing	428	0.366*	7.203*	10 574	0.255*	5.071*
Chemicals	671	0.214*	4.274*	24 871	0.147*	2.992*
Metals	1 023	0.178*	3.582*	38 440	0.047	1.067
Machinery	451	0.130*	2.655*	13 322	0.459*	9.003*
Electrical and electronic	511	0.168*	3.400*	18 320	0.051	1.129
Chennai						
Food and beverages	166	0.094	1.798	8 459	0.069	1.353
Textiles	1 446	0.135*	2.495*	107 266	0.039	0.694
Leather	406	0.058	1.161	17 919	0.087	1.666
Printing and publishing	402	0.230*	4.135*	9 361	0.141*	2.595*
Chemicals	512	0.126*	2.329*	17 920	0.132*	2.443*
Metals	695	0.184*	3.337*	29 076	0.006	0.266
Machinery	416	0.153*	2.799*	18 772	0.039	0.829
Electrical and electronic	359	0.127*	2.361*	21 018	−0.003	0.098

*Significant at the 0.05 level.
[a]I, global Moran index; Z, Z statistic.

Local clustering

In this section we identify the locations and other characteristics of the clusters. Recall from our earlier discussion on the measurement of clustering that local clusters may exist even when the system as a whole is not clustered (hence the distinction between 'global' and 'local' Moran indices). These results are reported in table 2 and in a series of maps (figures 1, 2, and 3, see over).

It is necessary to explain what is reported in figures 1–3 and in table 2. We began the analysis by calculating first-order local Morans for each of our eight industry categories, for each of our three cities, for factories and workers separately. In figure 1 the first pair of maps in the top left-hand corner shows the distribution of local Morans for the food and beverage sector in Mumbai; the map on the left in this pair shows the local Morans for factories; the map on the right shows the local Morans for workers. The strength of the clusters is determined by the statistical significance of the local Moran of each pin code. These Z-values for local Morans are mapped. A negative cluster has Z-values less than -1.65 (not shown). Z-values between -1.65 and 1.65 are not clustered. Values of Z between 1.65 and 1.95 are weakly clustered, and values greater than 1.95 are strongly clustered. Pin codes that show weak and strong evidence of positive clustering are shown in the maps by using different shading. In general, where local clustering does exist it does so at Z-values greater than 1.95. Last, the numbers of pin codes forming clusters and the numbers of factories and workers in their respective clusters were summed. These data are reported in table 2. Note the great variation in the results between cities, between industries, and within cities and within industries. Let us identify some of the observed patterns.

First, factories and workers in the same industry do not necessarily cluster in the same pin codes. In the general case there are some common pin codes and some unique pin codes for each industry. There are two types of exceptional cases: one where there is perfect overlap between factory and worker clusters (such as in printing and publishing in Mumbai, and in machinery in Chennai), the other in situations where there are no common pin codes (such as in food and beverages in Mumbai, where there are five unique clustered pin codes each for factories and for workers) or few common pin codes (such as for textiles or machinery in Kolkata). This variation is seen in every city. This is an important finding. It suggests that within the same industry, small-scale operations tend to cluster together (this is where the factories are seen to cluster), often, but not always, at separate locations from large-scale operations (where workers are seen to cluster). Later we discuss the implications of this finding.

Second, and related to the first point, it is difficult to discern whether factories or workers are more clustered in specific industries. Recall the argument that large factories rely on internal economies of scale for productivity gains, whereas smaller factories rely on external economies, at least some of which are derived from localization or clustering. Hence, in our data, we can expect factories (small-scale units) to cluster more and workers (large-scale units) to cluster less. If we use the percentage of factories or workers within clusters as a measure of the extent of clustering, this expectation is correct across cities in only the printing and publishing and the metals sectors. The opposite is true in the food and beverages sector, and, with the exception of Chennai (where the numbers are close), in the textiles sector. This may not be a significant issue as it is unclear that the measure used here is appropriate for comparing two very different types of units (factories and workers).

Third, each city has one or two industries that appear to be more clustered than others—for instance, the textiles and the electrical and electronic sectors in Mumbai, the leather and the food and beverage sectors in Kolkata, and the leather sector in Chennai.

Table 2. Concentration in clusters (data source: CSO, 1998 – 99).

	Factories			Workers			Pin codes			
	total number	in clusters		total number	in clusters		number in factory clusters	number in worker clusters	number in factory and worker clusters	degree of overlap[a]
		no.	%		no.	%				
Mumbai										
Food and beverages	274	79	28.8	16 992	7 791	45.9	5	5	0	low
Textiles	1 988	703	35.4	131 318	62 649	47.7	5	6	2	low
Leather	115	45	39.1	2 629	691	26.3	4	3	1	low
Printing and publishing	882	203	23.0	17 001	3 567	21.0	2	2	2	high
Chemicals	1 146	508	44.3	41 470	5 623	13.6	7	3	2	–
Metals	892	307	34.3	25 056	1 347	5.4	4	1	1	–
Machinery	767	185	24.1	40 808	0	0.0	3	0	0	–
Electrical and electronic	802	318	39.7	58 871	34 575	58.7	4	3	3	high
Kolkata										
Food and beverages	355	122	34.4	9 573	6 034	63.0	5	9	4	–
Textiles	355	120	33.8	90 621	38 271	42.2	6	7	1	low
Leather	283	209	73.9	6 149	4 393	71.4	6	8	6	high
Printing and publishing	428	209	48.8	10 574	4 155	39.3	10	9	9	high
Chemicals	671	224	33.4	24 871	10 053	40.4	7	9	7	high
Metals	1 023	357	34.9	38 440	8 166	21.2	4	3	2	–
Machinery	451	101	22.4	13 322	4 217	31.7	6	6	1	low
Electrical and electronic	511	170	33.3	18 320	5 283	28.8	9	6	4	–
Chennai										
Food and beverages	166	45	27.1	8 459	3 065	36.2	4	4	2	–
Textiles	1 446	313	21.6	107 266	22 447	20.9	7	3	2	–
Leather	406	204	50.2	17 919	9 932	55.4	4	6	4	high
Printing and publishing	402	161	40.0	9 361	2 534	27.1	8	5	5	high
Chemicals	512	137	26.8	17 920	5 333	29.8	6	5	2	low
Metals	695	240	34.5	29 076	5 755	19.8	4	3	2	–
Machinery	416	167	40.1	18 772	5 222	27.8	2	2	2	high
Electrical and electronic	359	157	43.7	21 018	1 065	5.1	4	1	1	–

[a] Only low and high degrees of overlap are indicated.

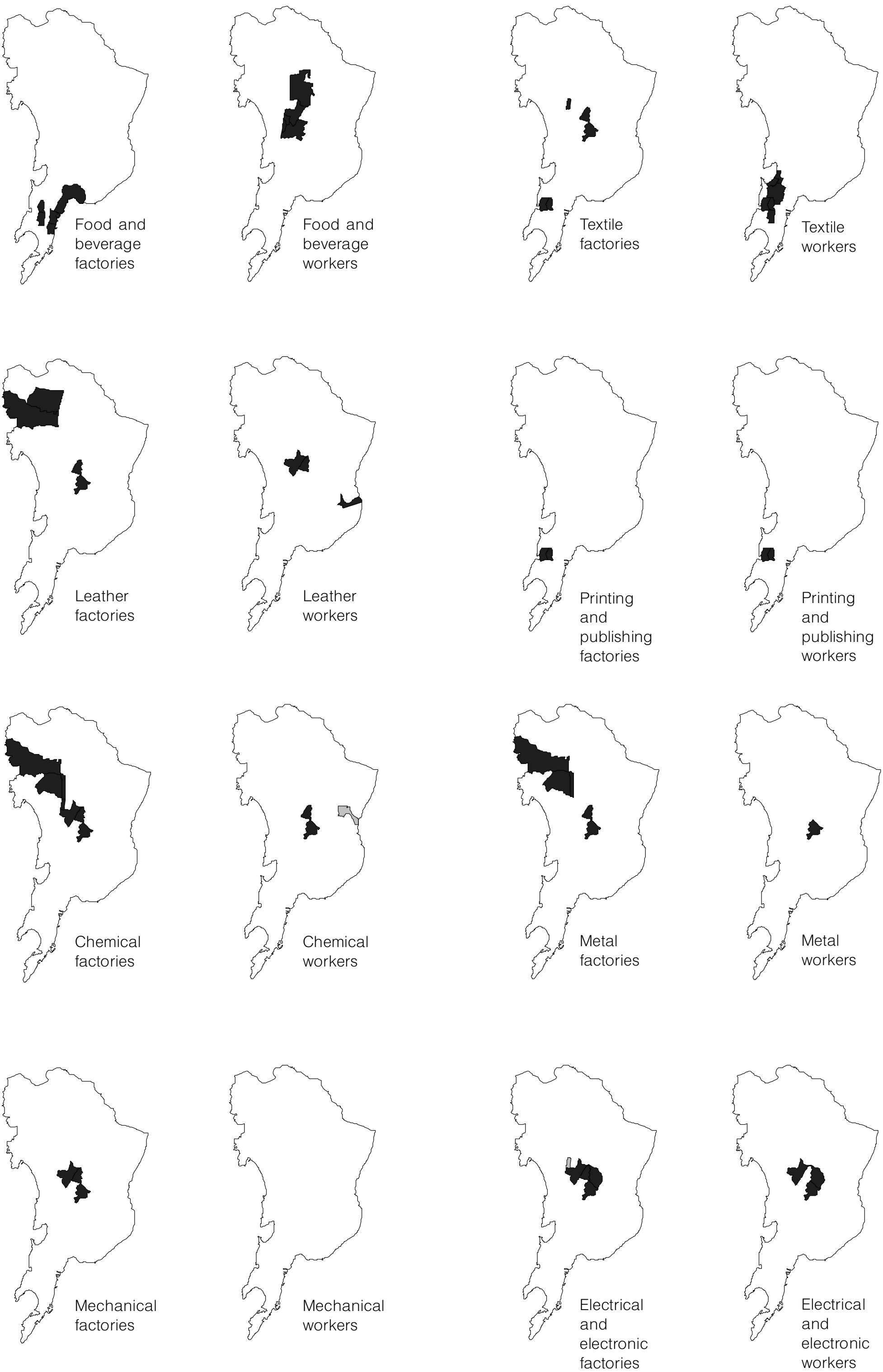

Figure 1. Industrial clusters in Mumbai (dark shading $Z > 1.95$, lighter shading $1.65 < Z \leqslant 1.95$).

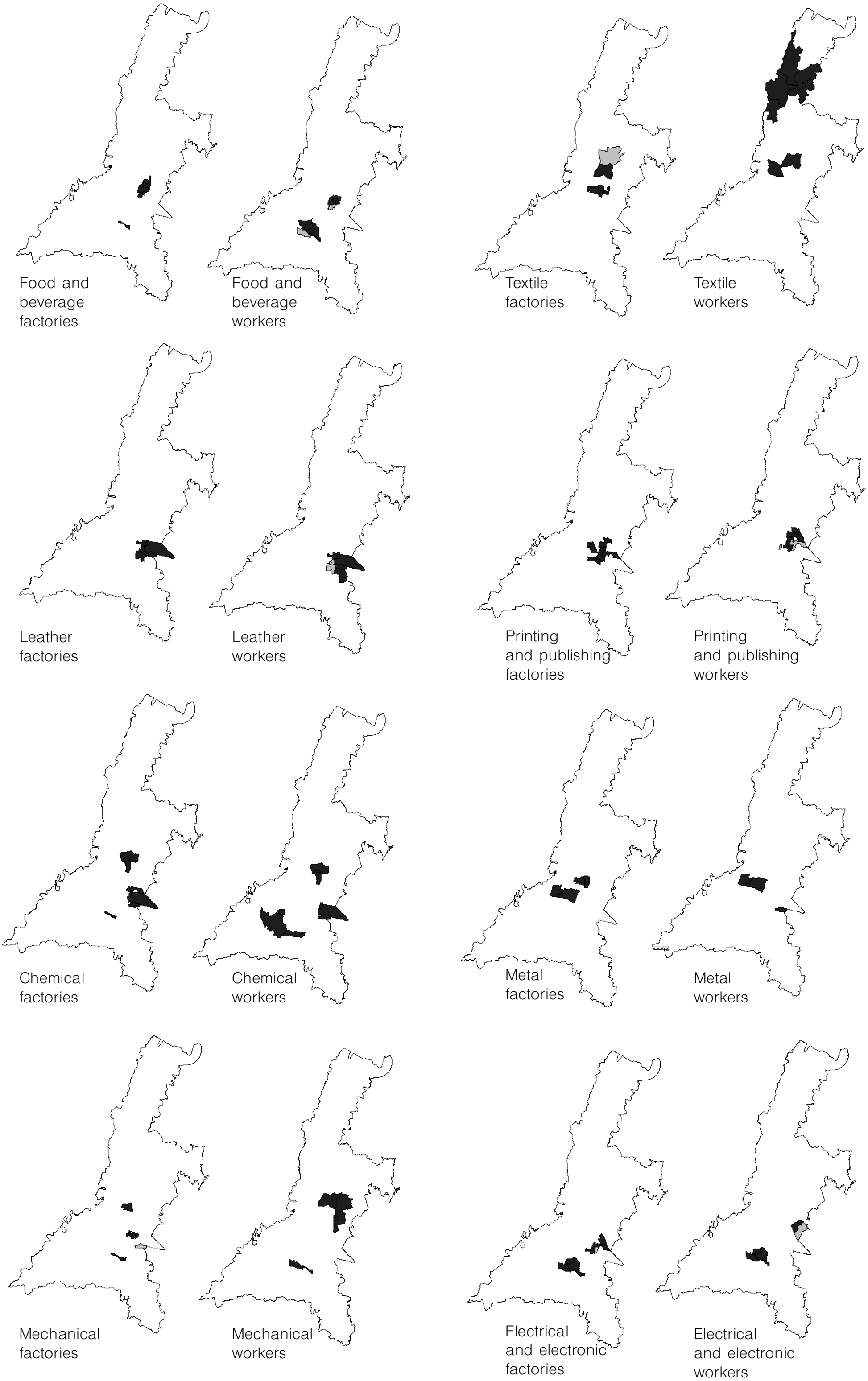

Figure 2. Industrial clusters in Kolkata (dark shading $Z > 1.95$, lighter shading $1.65 < Z \leqslant 1.95$).

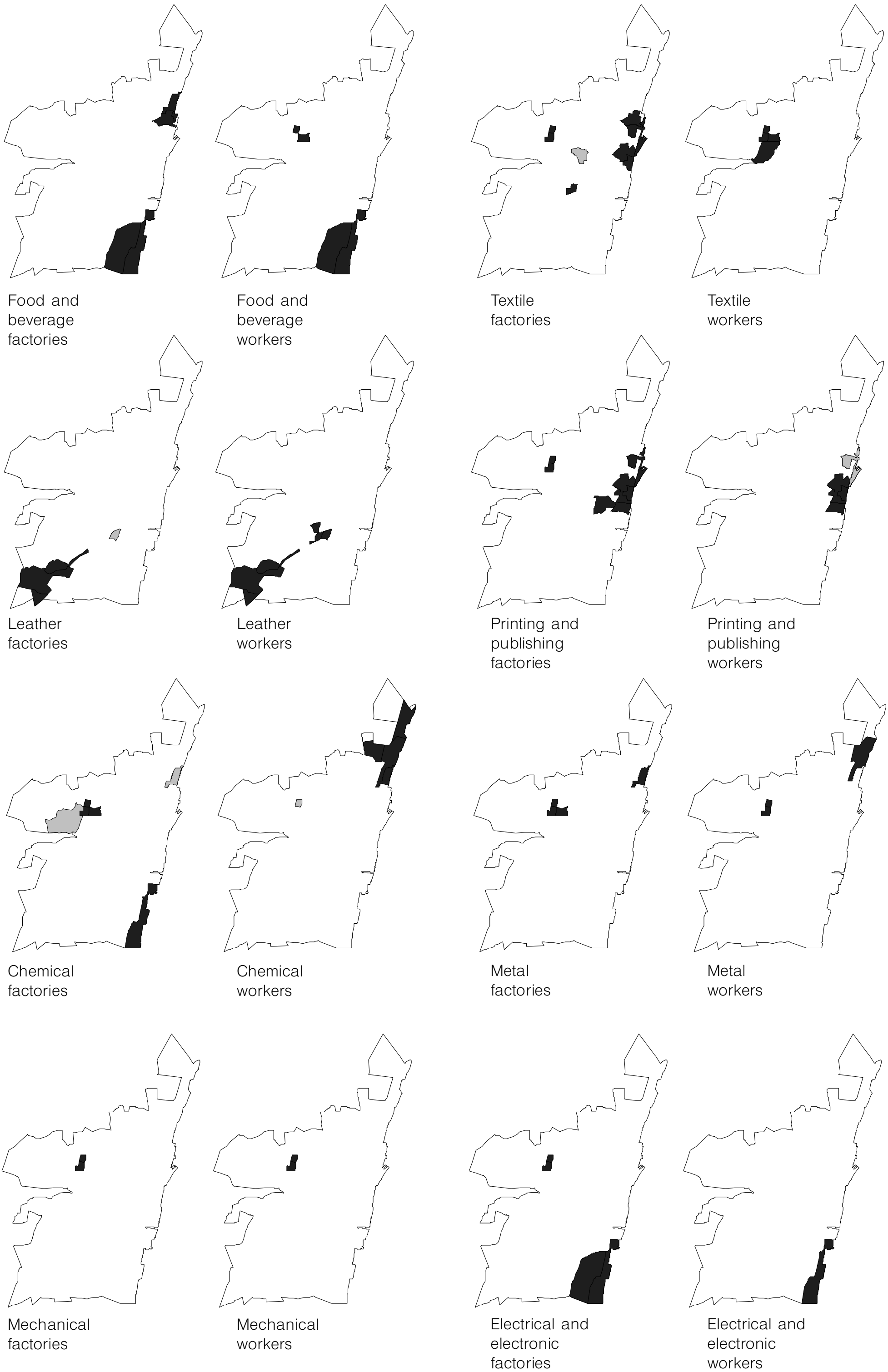

Figure 3. Industrial clusters in Chennai (dark shading $Z > 1.95$, lighter shading $1.65 < Z \leqslant 1.95$).

These are also the industries that generally have a significant overlap in terms of the locations of factory and worker clusters (with the exception of the textiles sector in Mumbai). Moreover, these are also the industries for which these cities have high location quotients (LQs) at the national level. [The LQ is a simple measure of regional concentration used in regional science. It calculates the ratio of the share of a given variable to the share of population. Here, LQ = 1 indicates that a pin code's share of a particular sector is equal to its share of all industry. If LQ = 3 it indicates that the pin code's share of that sector is three times its share of all industry.] The electrical and electronic sector in Mumbai has an LQ of 3.0; the LQ for the textile sector is 1.7. The LQ of the leather sector in Chennai is 4.0; in Kolkata it is 1.5. The food and beverages sector in Kolkata is an exception to this pattern. Another exception is the machinery sector in Mumbai; its LQ is 2.5, yet it is the only industry in any of our study cities to have absolutely no local clustering among workers (and one of the lowest levels of clustering among factories in all cities). Therefore, though this pattern cannot be generalized, there may be a causal relationship between very high levels of clustering and industry dominance at the national level. At this point, however, it is difficult to determine the direction of the causal arrow.

Fourth, the location of industry clusters generally appears to follow a couple of widely held principles: (1) polluting industries are located in fringe areas and (2) such polluting industries are located in proximity to each other. Consider the first principle. It seems obvious that any local regulatory agency will direct the location of polluting industry toward the urban fringe. The two most polluting sectors considered here are the chemicals sector and the leather sector. In all three cities these industries are located in fringe areas, and it appears from the maps that in the cases of Mumbai and Kolkata these two industries share common locations (see figures 1 and 2). Also note the particularly interesting case of the chemicals sector in Chennai, where factories are clustered in the southern extremity and in the far west, whereas workers (or large-scale factories) are located in the northern fringes of the metropolitan area.

Coclustering, colocation, and industrial districts

In this section we test for the existence of colocation or coclustering of different industry sectors. A brief discussion of theory, on why colocation or coclustering may be possible, is necessary before we discuss the tests and outcomes. It is useful at this point to distinguish between colocation and coclustering. Colocation occurs when industries from two sectors are present in the same neighborhood. Coclustering occurs if both industries that are colocated are related through economic (input–output, innovation, or labor-market) linkages.

Theory

The idea of the industrial district goes back to Marshall (1919), who suggested that small specialized firms would tend to cluster in space to derive external economies to offset the internal scale economies of large factories. Piore and Sabel (1984) argued that the late twentieth century had seen the arrival of a 'second industrial divide' where the vertically integrated organization of production characteristic of Fordist manufacturing was giving way to regional specialization organized around networks of small-scale producers. Geographers see this in terms of the need for flexible specialization in globalized production systems geared toward rapid changes in technology and the need to respond to shifting patterns of demand (Amin, 2002). Economists have focused on the specific productivity advantages provided by proximity. In the simplest terms, these localization economies (to be distinguished, as discussed earlier, from urbanization economies that accrue to all firms in an urban area) arise from two sources: local labor markets and knowledge spillovers.

Labor markets Do thick local labor markets create localization economies? There are two questions inherent in this issue: Are labor markets spatially segregated by skill or specialization? Do industries become spatially segregated in response to segregated labor markets? We need to distinguish between industries based on unskilled labor (for example, the leather industry) from those based on skilled labor (for example, the printing and publishing industry). Consider labor markets with unlimited supplies of unskilled labor operating in land markets where land rents decline with proximity to industry. In other words, unskilled labor is available in plenty and can locate anywhere, but is likely to locate near factories where rents are lowest. Therefore, unskilled labor is likely to colocate with industrial clusters. Does this generate localization economies for those industries that are reliant on unskilled labor? If it does we are likely to see industries based on unskilled labor oriented more toward locating in industrial clusters with other such industries that do not necessarily share other characteristics. These clusters will not include industries based on skilled labor. If these (skill-based) industries do cluster, they do not do so for labor-market localization economies, as their critical workers, the high-skill high-wage labor, will not colocate with industry. They do so for other reasons (discussed below). In summary, labor-market localization economies, if any exist, are likely to be industry specific and inversely related to the proportion of skilled labor in a given industry.

Knowledge spillovers These are of two kinds—technology spillovers through informal interaction, and information spillovers on market agents. Technology spillovers are irrelevant in low-technology firms and in industries in mature stages of the product cycle. The vast majority of manufacturing industry in India belongs in this category. Hence, for these industries we can argue that there are negligible localization economies from technology spillovers. Information spillovers on market agents such as buyers and suppliers, however, are more likely to provide localization economies. Firms of all sizes (except perhaps the very largest vertically integrated firms) rely on dense buyer – supplier networks. Firms benefit from having access to local buyers and suppliers, and knowledge pooling on buyer – supplier behavior is likely to eliminate inefficient agents.

Conclusion Hence, theory suggests the existence of two types of industrial districts: *labor-sharing* industrial districts that depend on the local availability of low-skill labor, and *buyer – supplier-linked* industrial districts, where industries that have market interactions with each other benefit from colocation. We do not have spatially disaggregated wage data to test explicitly for the existence of labor-sharing districts. We are, however, able to establish what the expected buyer – supplier links are between our industry groups by assuming that the input – output (IO) links at the national level are replicated at the local level. The national IO tables are available. We assume that similar IO linkages exist at the level of the metropolis and argue that industries with strong IO linkages are likely to colocate.

Evidence

Before we examine the evidence on buyer – supplier linkages, let us first examine the data on colocation by industry for all sectors. The correlation coefficients for factories and workers for all eight industry groups are reported in table 3. The numbers here are quite remarkable. There is strong evidence that industry groups colocate at the pin code level, especially in the case of factories or small-scale units. In Mumbai, for instance, factories for every industry group are seen to have a statistically significant correlation with every other industry group. In general, the correlation coefficients are very high: 0.93 between machinery and chemicals, 0.91 between chemicals and metals and so on.

Table 3. Correlation coefficients for industry pairs (data source: CSO, 1998 – 99).

	Food and beverages	Textiles	Leather	Printing and publishing	Chemicals	Metals	Machinery	Electrical and electronic
Mumbai								
Food and beverages	1.00	0.39**	0.31**	0.39**	0.34**	0.40**	0.40**	0.25**
Textiles	0.10	1.00	0.58**	0.79**	0.77**	0.73**	0.77**	0.63**
Leather	0.20**	0.23	1.00	0.43**	0.77**	0.63**	0.75**	0.63**
Printing and publishing	0.16	0.74**	0.35**	1.00	0.49**	0.44**	0.51**	0.45**
Chemicals	0.25**	0.37**	0.42**	0.33**	1.00	0.91**	0.93**	0.81**
Metals	0.16	0.26**	0.18	0.25**	0.56**	1.00	0.87**	0.70**
Machinery	0.09	0.17	0.15	0.07	0.34**	0.15	1.00	0.78**
Electrical and electronic	0.27**	0.07	0.16	0.10	0.21**	0.25**	0.06	1.00
Kolkata								
Food and beverages	1.00	0.31**	0.09	0.46**	0.43**	0.33**	0.46**	0.40**
Textiles	0.07	1.00	0.04	0.36**	0.64**	0.33**	0.51**	0.51**
Leather	0.02	−0.05	1.00	0.03	0.49**	0.07	0.12	0.10
Printing and Publishing	0.12	0.03	0.08	1.00	0.34**	0.15	0.37**	0.32**
Chemicals	0.17	0.19**	0.33**	0.10	1.00	0.36**	0.60**	0.72**
Metals	0.14	0.09	0.14	0.07	0.25**	1.00	0.76**	0.33**
Machinery	0.21**	−0.03	0.04	0.26**	0.22**	0.41**	1.00	0.68**
Electrical and electronic	0.25**	0.06	0.02	0.08	0.37**	0.08	0.30**	1.00
Chennai								
Food and beverages	1.00	0.30**	0.02	0.47**	0.69**	0.59**	0.52**	0.62**
Textiles	0.15	1.00	0.29**	0.61**	0.51**	0.40**	0.41**	0.40**
Leather	−0.01	0.35**	1.00	0.10	0.16	0.11	0.14	0.15
Printing and publishing	0.25**	0.32**	0.17	1.00	0.58**	0.56**	0.57**	0.46**
Chemicals	0.27**	0.34**	0.23**	0.33**	1.00	0.87**	0.84**	0.79**
Metals	0.22**	0.31**	0.13	0.63**	0.33**	1.00	0.95**	0.71**
Machinery	0.41**	0.39**	0.09	0.31**	0.41**	0.34**	1.00	0.76**
Electrical and electronic	0.23**	0.50**	0.27**	0.26**	0.40**	0.23**	0.27**	1.00

**Significant at the 0.01 level.
Note: factory data are above the diagonal; worker data are below the diagonal.

In the case of workers, however, the correlations are not as high, and fewer are statistically significant. In Chennai, the pattern is even more pronounced. With the exception of the leather sector (which has a moderate but significant correlation with only the textiles sector) the correlation coefficients of every other pair of industries is significant, and, as in the case of Mumbai, often very high. The leather sector is anomalous in Kolkata as well. Its only significant correlation in that metropolis is with the chemicals sector. But Kolkata itself is somewhat of an anomaly compared with Mumbai and Chennai, at least in terms of the colocation of workers. Factories in Kolkata, with the exception of the leather sector, are generally colocated; however, workers in at least four industries—food and beverages, textiles, leather, and printing and publishing—are generally not correlated with workers in other industries. These findings provide further evidence that small units are more likely to cluster than are large units.

It is one thing for industries to be colocated, it is another for them to be coclustered. If the reasoning on buyer–supplier clusters outlined above is correct we can expect: (a) that industries that have strong IO links will cocluster; that is, they will form clusters in the same or proximate pin codes, and (b) industries that do not share buyer–supplier linkages will not cocluster; if they do it will be for labor sharing, and therefore they will share similar labor profiles (low skill with low skill, or high skill with high skill). In order to test this hypothesis we conducted coclustering tests on four industry pairs; two of these pairs have strong IO linkages (the two highest among our eight industry groups)—metals and machinery, with an IO factor of 34.95, and metals and electrical and electronics, with an IO factor of 25.75. The two other pairs have virtually no IO linkages—food and beverages and electrical and electronics with an IO factor of 0.07 and very different labor profiles (the former is low skill, the latter is high skill); and textiles and metals, with an IO factor of 0.41 and potentially similar labor profiles. We combine the proportion (rather than raw) data on factories and workers for these four industry pairs by pairs, and conduct tests of global clustering (Moran's I) and local clustering.

The results are mixed. Consider the results of the global clustering test first (shown in table 4). In Mumbai, among the strong IO pairs, only factories in the metals and electrical and electronics combination are clustered, whereas the food and electrical and electronics combination, which shares neither IO links nor similar labor profiles, shows clustering for factories and workers. In Kolkata, all four pairs are clustered, for factories and workers (more for factories than for workers). In Chennai, as in Mumbai, factories in the metals and electrical and electronics combination are clustered; but factories in both weak IO combinations are clustered. The strongest IO pair—metals and machinery—is not only not clustered in Mumbai and Chennai, but the value of Moran's I is negative half the time.

Next, for further evidence on colocation, we look at a final set of data: the share of industry in the top pin codes. In table 5 (see over) we list these data by metropolis, for all industry and by sector. (Note that in a given metropolis the top ten or top twenty pin codes are not separately identified for each sector. What is reported is the share of each sector in the top overall pin codes.) The results are not surprising in the context of what we have presented earlier, but they are quite effective in making the point that industry is concentrated in a handful of pin codes in all three cities, and that most sectors are heavily represented in these top pin codes. In Mumbai the top ten pin codes include close to 55% of all factories and workers, and the top twenty pin codes include over 76% of all factories and nearly 80% of all workers. In Chennai and Kolkata the proportions are progressively smaller. This decline is probably a function of the fact that the total number of pin codes in Kolkata (133) and Chennai (108) are higher than in Mumbai (94). In general, regardless of the number of pin codes in a metropolis, the top 10% of the pin codes include close to 50% of all factories and workers.

Table 4. Indices of coclustering in selected industry pairs (data source: CSO, 1998–99).

	National IO factor	Factories			Workers		
		total number	*I*	*Z* of *I*	total number	*I*	*Z* of *I*
Mumbai							
Strong IO links							
Metals and machinery	34.95	1 661	0.082	1.496	65 864	−0.052	0.668
Metals and electrical and electronics	25.75	1 696	0.161*	2.774*	83 927	0.105	1.876
Weak IO links							
Food and beverages and electrical and electronics	0.07	1 076	0.184*	3.145*	75 863	0.294*	4.936*
Textiles and metals	0.41	2 882	0.080	1.473	156 374	0.045	0.904
Kolkata							
Strong IO links							
Metals and machinery	34.95	1 474	0.222*	4.429*	51 762	0.121*	2.490*
Metals and electrical and electronics	25.75	1 534	0.211*	4.222*	56 760	0.109*	2.249*
Weak IO links							
Food and beverages and electrical and electronics	0.07	866	0.178*	3.579*	27 893	0.185*	3.730*
Textiles and metals	0.41	1 378	0.270*	5.349*	129 061	0.181*	3.644*
Chennai							
Strong IO links							
Metals and machinery	34.95	1 111	−0.079	1.060	47 848	0.029	0.673
Metals and electrical and electronics	25.75	1 054	0.149*	2.738*	50 094	−0.004	0.087
Weak IO links							
Food and beverages and electrical and electronics	0.07	525	0.126*	2.332*	29 477	0.041	0.873
Textiles and metals	0.41	2 141	0.161*	2.948*	136 342	0.026	0.623

*Significant at the 0.05 level.
Note: *I*, global Moran index; *Z*, *Z* statistic; IO, input–output.

Table 5. Industry concentration in top districts (data source: CSO, 1998, 99).

	Number of units	Share of metropolitan total (%)	Sectoral share (%)							
			food and beverages	textiles	leather	printing and publishing	chemicals	metals	machinery	electrical and electronics
Mumbai										
Factories										
top ten	3744	55.20	23.73	57.75	56.52	45.69	57.77	60.40	56.32	53.49
next ten	1447	21.34	25.18	22.69	19.13	16.78	21.73	20.69	23.99	17.33
total[a]	5191	76.54	48.91	80.43	75.65	62.47	79.49	81.10	80.31	70.82
Workers										
top ten	176701	53.19	28.05	51.18	20.27	35.45	34.17	50.00	62.75	77.91
next ten	86411	26.01	15.70	32.58	51.16	23.27	29.80	24.15	25.16	11.85
total[a]	263112	79.20	43.75	83.76	71.44	58.72	63.97	74.15	87.91	89.76
Kolkata										
Factories										
top ten	1610	36.38	24.35	38.64	59.25	26.44	29.70	46.22	35.54	26.37
next ten	787	17.79	25.13	13.05	21.57	16.78	18.08	13.72	18.88	22.41
total[a]	2397	54.17	49.48	51.70	80.82	43.22	47.78	59.93	54.42	48.78
Workers										
top ten	77588	33.73	11.19	57.66	0.42	7.00	19.81	19.02	9.68	25.57
next ten	49117	21.35	30.47	23.14	42.00	15.09	13.98	23.82	13.33	13.40
total[a]	126705	55.08	41.66	80.80	42.42	22.09	33.79	42.84	23.01	38.97
Chennai										
Factories										
top ten	1872	42.53	33.74	27.25	53.20	28.36	40.62	51.22	69.95	66.02
next ten	792	17.99	21.08	23.03	6.16	33.83	18.16	17.12	5.53	7.80
total[a]	2664	60.52	54.82	50.28	59.36	62.19	58.79	68.35	75.48	73.82
Workers										
top ten	104902	45.65	23.34	40.60	53.58	35.54	42.91	51.12	44.82	73.68
next ten	45364	19.74	29.32	20.44	14.09	14.56	15.17	16.37	38.04	11.66
total[a]	150266	65.39	52.66	61.04	67.66	50.10	58.08	67.49	82.85	85.34

[a] Top twenty

Summary of the findings

First, in Indian metropolises industry is generally clustered—the evidence for clustering is found at the level of the metropolis (using the global Moran statistic) and at the level of the pin code (using local Moran statistics and maps). At the sectoral level there are extremes: in one case there are no clusters at all; at the other extreme there are instances where 70% or more of workers or factories are concentrated within six to eight pin codes (see table 2).

Second, the clusters are of two types: (1) where factories and workers are both clustered in the same region within a metropolis, and (2) where factories are clustered at locations separate from worker clusters (that is, separate clusters of small-scale operations and large scale operations). Both patterns are equally common. In general, factories (small units) are more clustered than are workers (large units).

Third, a small number of pin codes account for a very large proportion of all industry, both in terms of factories and in terms of workers. As a result, the extent of industry colocation is very high. However, the expected relationships between industrial sectors—whereby industries with strong IO linkages are expected to colocate, and industries using similar labor profiles are expected to colocate—are not found. On the contrary, we see several examples of counterintuitive colocations.

Last, some industries have distinct locational properties. For example, where the leather industry is significant (as in Kolkata and Chennai) it is located on the urban fringe and is not co-located with other industries except the chemicals industry. The printing and publishing industry is located near the urban core in all three cities. Both industries are also highly clustered. The textile industry, the largest industry, is marked, in contrast, by separate factory and worker clusters, where the factory cluster is closer to the city center than is the worker cluster (except in Mumbai).

An explanatory framework

How can these patterns be explained? To begin with, we suggest that the conventional wisdom—according to which, localization economies drive cluster formation—may be limited in its explanatory power. There is little evidence in support of the processes of localization, either via local labor markets or via local buyer–supplier networks. It is possible to argue that these findings are artifacts of the method of local cluster identification. That is, had we used other methods that would have enlarged the definition of 'local' beyond the one used here (a pin code and its adjacent neighbors), there would have been stronger evidence in support of localization economies driving cluster formation. That, however, is unlikely. First, at a preliminary stage in our investigations we did indeed use larger definitions of 'local', with very similar results. Second, it may be possible to enlarge the definition of 'local' until it is meaningless or large enough to encompass a significant portion of the metropolitan area. At that point it becomes difficult to distinguish localization economies from urbanization economies.

One can also argue that localization economies arise primarily as a result of interfirm trade or technological interaction *within* an industrial sector as defined in this analysis. That is, the sectors have been so broadly aggregated that most trade and technical exchange takes place *within* rather than *between* sectors. For many sectors it is difficult to make that case, as according to the national IO data, interfirm trade within most of these sectors is not very high. For instance, within the food and beverages, chemicals, and machinery sectors the intrasector IO coefficients are under 5.5. The other sectors, where intrasector IO links are stronger, are also dominated by old firms that are not close to the technological frontier (Lall and Rodrigo, 2001); as a result, there is little likelihood of intrasector technical exchange.

Therefore, we must conclude that *buyer – supplier networks and labor pools are metropolis-wide rather than localized*; hence, the debate on the relative strength of localization and urbanization economies for industrial location decisions should be resolved firmly in favor of urbanization.

Looking beyond localization economies, we suggest that firm locations are guided by a complex set of factors that often rule out most spaces within metropolitan areas. These factors include accidents of history, metropolitan expansion, industry characteristics, and, above all, state regulations (especially ones that affect the land market). As a result, we see the unplanned or planned evolution of *mixed industrial districts,* which include a variety of related and unrelated industry sectors, at leapfrogging locations within metropolitan areas. We suggest that a historical framework may explain the observed patterns of industry location in Indian metropolises. This framework is speculative (we do not have the temporal information needed to make a stronger claim), but it is able to account for many of the expected and unexpected findings.

First, at some specific historic point a pioneer industrial unit in a specific industrial sector makes a location decision within the metropolis. The driving force behind that decision (not the location decision, but the decision to start a new industry) may have one or several motivations: nationalism (as in the case of the first textile mill in 1854 in Mumbai), war (as in the case of the first leather units in Kolkata, created to outfit saddles for the Imperial army during the First World War), bureaucratization and the spread of literacy (creating the need for printed matter), etc. This first factory did not rely on localization economies to boost productivity but probably relied on urbanization economies, at least in terms of providing market access and a pool of labor. This unit located in what was then the urban fringe, beyond dense population settlements, but close enough for workers to reach the plant. Following the convention of the time, this unit was large, relying on internal economies of scale to reduce costs.

A cluster of firms in the same industry began forming around this original location. It is difficult to determine whether these subsequent location decisions were the result of localization advantages arising from labor pooling, advantages derived from shared infrastructure such as railheads, or state regulations that directed new firms to this location. It will be necessary to conduct archival research for more concrete judgments. It is very likely that at some point industries in this cluster began to derive benefits from labor pooling. Two factors should be considered here. First, these were not the high-technology industries of the time; they were industries in the late stages of the product cycle (Vernon, 1966), reliant on unskilled and semiskilled labor. Second, the location of these industries began influencing the land market around them; because of the local environmental impacts of these industries, the only bidders for the proximate land were other industries or low-wage labor. In other words, this subregion became an industrial district, with large-scale factories, and slums for unskilled workers. As the city continued to grow beyond this industrial district most new industry was directed to this district by state regulations. The single-purpose industrial district became a mixed industrial district.

Where was even newer industry to locate? With all the land in the original industrial district in use, and with state regulations that forbade the conversion of any land with housing (slums, tenements, or middle-class residences) to industrial use, new industries now sought new locations on what was then the urban fringe. The role of the state in discouraging land-use change turns out to be a critical influence on industrial location (as we shall see below). The cycle of industrial district formation began again, this time with more active involvement of the postcolonial state, which assisted with land acquisition (often the most difficult aspect of industrial location in urban areas) and provided physical infrastructure.

When this cycle was complete (that is, when the point was reached when there was no more land for new factories) new industrial units leapfrogged over the residential communities that had grown in the interim to new locations at the urban fringe. This is the current stage. Now there is even more active involvement of the state, which sets up export processing zones, free-trade zones, technology parks, industrial parks, and so on to entice new industrial units. (Consider the names of some of the most active pin codes: in Mumbai, Chakala and SEEPZ, both Maharashtra Industrial Development Corporation centers; in Chennai, Ambattur Industrial Estate and SIDCO Industrial Estate, both set up by the state of Tamil Nadu.)

This stylized framework explains a number of observations—some regularities, some irregularities—listed in the preceding section. In this framework, the exact location of specific industries has to be understood in terms of the functions of these industries. For instance, the printing and publishing industry remains located close to the central business district (CBD), which is its principal market area. The leather industry, which pollutes both the air (with the smell of animal hides) and the groundwater, and therefore cannot even colocate with other industries (not to mention residential areas), remains at the urban fringe—when the fringe moves beyond or envelopes the leather cluster, the state compels the cluster to move further out. This is exactly what has happened in Chennai and is happening in Kolkata.

This framework also explains some anomalies, such as the location of a large textile cluster near the Mumbai CBD. This cluster should not exist, for the land there is too valuable to remain devoted to an industry with outmoded technology and very low value addition, yet it remains a textile cluster, with virtually closed factories, state takeover (or nationalization) of 'sick' units, and job losses that mount by the year, because the state will not permit the conversion of this industrial land to commercial or residential use (D'Monte, 2000).

The other significant anomaly comes from the relative locations of factory and worker clusters. It is not unusual to see these clusters form separately, but the expectation is that worker clusters (which are clusters of large factories), which require more undivided land, will locate on the least expensive land, furthest from the center. Yet when there is no exit policy (that is, factories are not allowed to close and factory land cannot easily be transferred) it is possible to see clusters of large-scale units near the center of the city. (Recall that many of the early industrial units, following the Fordist principles of the times, were large units.) And, most important, this framework explains why colocation is so common yet theoretical expectations on coclustering are not realized in practice.

Conclusions

Our principal conclusion is that land-market rigidities created by state policy rather than market opportunities in the form of localization economies have paramount importance in guiding intrametropolitan industrial locations. The key variable is land-use policy: segregationist or environmental policies that isolate polluting industries such as those of leather and chemicals; the absence of exit policies and land-use change policies, which keep obsolete large factories open near the CBD; and activist industry attraction and promotion policies that lead to the creation of new industrial zones that include a variety of industry sectors. This can be a serious problem in developing location theory, which assumes the operation of unhindered market forces, especially in the land market. The evidence suggests that intrametropolitan industrial location decisions are significantly influenced by state regulations. Does this mean that firms cannot utilize spatial economies? Does state intervention neutralize and perhaps even negate the spatial economies that would be possible in unfettered land markets?

In theory, this possibility certainly exists. However, the fact is that industries do end up being clustered. In fact, by reducing the location choices for firms, state policies probably help in cluster formation. If spatial economies in production arise from being clustered, regardless of whether the clustering came about as a result of market forces or state action, then we must conclude that these firms can enjoy at least some spatial economies. That is, state intervention influences industry location in significant ways but, having led to industrial clustering, should have relatively little impact on the generation of external economies during the production phase.

We must remember that there is a significant disconnection between the 'revealed preference' of firms (expressed in clustering) and their 'declared preference' in survey-based studies that show that there is a substantial random element in the choice of location: personal reasons, chance, and opportunity (especially that of finding a good site) are given as explanations almost half the time; proximity to other similar firms is not a serious factor (see Calzonetti and Walker, 1991; Mueller and Morgan, 1962). Therefore, it may be useful to reconsider some aspects of location theory from the perspective of what institutional economists call habit (Hodgson, 1998) and social theorists call *habitus* (Bourdieu, 2002). These theories suggest that most decisions are made on the basis of perceptions and are characterized by imitation, inertia, and cumulative causation. These are nevertheless efficient decisions because imitation and cumulative causation reduce risk and the cost of decisionmaking. When land markets are constrained, it makes sense to locate where other firms locate. The benefits of proximity may follow during the production phase, but during the location decision phase firms may be more interested in risk reduction than in seeking localization economies.

Acknowledgements. The findings reported in this paper are those of the authors alone and should not be attributed to the World Bank, its executive directors, or the countries they represent. This paper is part of a larger research effort on understanding the spatial dimensions of industry location, "The Economic Geography and Political Economy of Industrial Location", funded by World Bank research grant 77960. We are grateful to the Central Statistical Organization (CSO) of the Government of India for making the firm-level data available for this study. We would like to thank three anonymous referees, participants of the 2003 UNU/WIDER Tokyo conference on "Spatial Inequality in Asia", and the RSAI 2003 conference in Philadelphia for helpful comments and suggestions.

References

Amin A, 2002, "Industrial districts", in *A Companion to Economic Geography* Eds E Sheppard, T J Barnes (Blackwell, Oxford) pp 149 – 168

Anselin L, 1995, "Local indicators of spatial association: LISA" *Geographical Analysis* **27** 93 – 115

Bourdieu P, 2002, "Structures, habitus, practices", in *Contemporary Sociological Theory* Eds C Calhoun, J Gerteis, J Moody, S P Faff, I Virk (Blackwell, Oxford) pp 276 – 288

Calzonetti F J, Walker R T, 1991, "Factors affecting industrial location decisions: a survey approach", in *Industry Location and Public Policy* Eds H W Herzog, A M Schlottmann (University of Tennessee Press, Knoxville, TN) pp 221 – 240

Chakravorty S, 2000, "How does structural reform affect regional development? resolving contradictory theory with evidence from India" *Economic Geography* **76** 367 – 394

CSO, 1998 – 99, "Annual Survey of Industries Sampling Frame", Central Statistical Organization, Government of India, New Delhi

D'Monte D, 2000 *Ripping the Fabric: The Decline of Mumbai and its Mills* (Oxford University Press, New Delhi)

Ellison G, Glaeser E L, 1997, "Geographic concentration in US manufacturing industries: a dartboard approach" *Journal of Political Economy* **105** 889 – 927

Fujita M, Krugman P, Venables A, 1999 *The Spatial Economy: Cities, Regions, and International Trade* (MIT Press, Cambridge, MA)

Getis A, Ord J K, 1992, "The analysis of spatial association by use of distance statistics" *Geographical Analysis* **24** 189 – 206

Greenhut J, Greenhut M L, 1975, "Spatial price discrimination, competition and locational effects" *Economica* **42** 401 – 419
Henderson J V, Kuncoro A, 1996, "Industrial centralization in Indonesia" *World Bank Economic Review* **10** 513 – 540
Hodgson G M, 1998, "The approach of institutional economics" *Journal of Economic Literature* **36** 166 – 192
Hotelling H, 1929, "Stability in competition" *Economic Journal* **39** 41 – 57
Isard W, 1956 *Location and Space-economy* (MIT Press, Cambridge, MA)
Krugman P, 1991 *Geography and Trade* (MIT Press, Cambridge, MA)
Krugman P, 1996, "Urban concentration: the role of increasing returns and transport costs" *International Regional Science Review* **19** 5 – 30
Lall S V, Rodrigo C, 2001, "Perspectives on the sources of heterogeneity in Indian industry" *World Development* **29** 2127 – 2143
Lall S V, Shalizi Z, Deichmann U, forthcoming, "Agglomeration economies and productivity in Indian industry" *Journal of Development Economics*; also WP 2663, World Bank Policy Research, Washington, DC
Lee K S, 1989 *The Location of Jobs in a Developing Metropolis: Patterns of Growth in Bogotá and Cali, Colombia* (Oxford University Press, New York)
Losch A, 1956 *The Economics of Location* (Yale University Press, New Haven, CT)
Marshall A, 1919 *Industry and Trade* (Macmillan, London)
Martin R, Sunley P, 2003, "Deconstructing clusters: chaotic concept or policy panacea?" *Journal of Economic Geography* **3** 5 – 35
ML Infomap, 1998 *IndiaMap* (ML Infomap Pvt Ltd, New Delhi)
Mueller E, Morgan J N, 1962, "Location decisions of manufacturers" papers and proceedings of the Seventy-Fourth Annual Meeting of the American Economic Association *American Economic Review* **52** 204 – 217
Nadvi K, Schmitz H, 1999, "Industrial clusters in developing countries (special topic)" *World Development* **27** 1503 – 1704
Nayar B R, 1998, "Business and India's economic policy reforms" *Economic and Political Weekly* **33** 2453 – 2468
Parr J B, 2002, "Agglomeration economies: ambiguities and confusions" *Environment and Planning A* **34** 717 – 731
Piore M J, Sabel C F, 1984 *The Second Industrial Divide: Possibilities for Prosperity* (Basic Books, New York)
Porter M, 1990 *The Competitive Advantage of Nations* (Macmillan, London)
Porter M, 1996, "Competitive advantage, agglomeration economies, and regional policy" *International Regional Science Review* **19** 160 – 176
Ratcliffe J, 2004, "Geocoding crime and a first estimate of a minimum acceptable hit rate" *International Journal of Geographical Information Science* **18** 61 – 72
Vernon R, 1966, "International investment and international trade in the product cycle" *Quarterly Journal of Economics* **80** 90 – 207
Walker R A, 2000, "The geography of production", in *A Companion to Economic Geography* Eds E Sheppard, T J Barnes (Blackwell, Oxford) pp 113 – 132
Weber A, 1929 *Theory of the Location of Industries* (University of Chicago Press, Chicago, IL)
White M J, 1983, "The measurement of spatial segregation" *American Journal of Sociology* **88** 1008 – 1019
World Bank, 1999 *Entering the 21st century: World Development Report 1999/2000* (World Bank, Washington, DC)