WPS 3072

POLICY RESEARCH WORKING PAPER   3072

Diversity Matters

The Economic Geography of Industry Location in India

Somik V. Lall
Jun Koo
Sanjoy Chakravorty

Abstract

How does economic geography influence industrial production and thereby affect industrial location decisions and the spatial distribution of development? For manufacturing industry, what are the externalities that matter, and to what extent? Are these externalities spatially localized? Lall, Koo, and Chakravorty answer these questions by analyzing the influence of economic geography on the cost structure of manufacturing firms by firm size for eight industry sectors in India. The economic geography factors include market access and local and urban externalities—which are concentrations of own-industry firms, concentrations of buyer-supplier links, and industrial diversity at the district (local) level. The authors find that industrial diversity is the only economic geography variable that has a significant, consistent, and substantial cost-reducing effect for firms, particularly small firms. This finding calls into question the fundamental assumptions regarding localization economies and raises further concerns on the industrial development prospects of lagging regions in developing countries.

Produced by Partnerships, Capacity Building, and Outreach

Diversity Matters [*]

The Economic Geography of Industry Location in India

Somik V. Lall
Development Research Group
The World Bank, Washington, DC, USA
slall1@worldbank.org

Jun Koo
Development Research Group
The World Bank, Washington, DC, USA
kkyojun@worldbank.org

Sanjoy Chakravorty
Department of Geography and Urban Studies
Temple University, Philadelphia, PA, USA
sanjoy@temple.edu

Diversity Matters

Where do different industries locate? What factors influence the spatial distribution of economic activity within countries? Finding answers to these questions is important for understanding the development potential of sub national regions. This is particularly important for developing countries as they have relatively lower levels of overall investment and economic activity is concentrated in one or a few growth centers. Thus, regions that do not attract dynamic industries are not only characterized by low productivity, but also by lower relative incomes and standards of living. These questions on industry location and their implications are not new. Examining the locational aspects of economic activity has long been of interest to geographers, planners, and regional scientists (Weber, 1929; Lösch, 1956; Hotelling, 1929; Greenhut and Greenhut, 1975, Isard 1956, Von Thunen, 1966). However, analytic difficulties in modeling increasing returns to scale marginalized the analysis of geographic aspects in mainstream economic analysis (Krugman 1991a). Recent research on externalities, increasing returns to scale, and imperfect spatial competition (Dixit and Stiglitz 1977; Fujita, et al. 1999; Krugman 1991b) has led to renewed interest in analyzing the spatial organization of economic activity. This is especially true in the case of geographic concentration or clustering.

Models in the 'New Economic Geography' literature (see review in Fujita, Krugman, and Venables, 1999) allow us to move from the question 'Where will manufacturing concentrate (if it does)?' to the question 'What manufacturing will concentrate where?' These insightful theoretical models provide, for the most part, renewed analytical support for the "cumulative causation" arguments made in earlier decades on the core-periphery relationship, on agglomeration economies, and on industrial clustering. We are interested in finding empirical answers to these (very old) questions, and to go beyond, to ask, "What manufacturing will locate where and *why*"?

1

To understand the process of industrial location and concentration, it is important to first analyze the location decisions of firms in particular industries. The location decision of the individual firm may be influenced by several factors. These include (a) availability of infrastructure, and the external economies provided by localization and urbanization, i.e., the "economic geography", (b) local wages, taxes, subsidies, and incentives, i.e., the "political economy", and (c) history, being "accidental". Here we focus on the economic geography characteristics. We develop and estimate an economic model to assess the impacts of region specific characteristics on location choices of firms in well defined industries. For the empirical application, we use micro level establishment data for Indian industry to examine the contribution of regional characteristics on location choices. Our concept of regional characteristics extends beyond its natural geography. Rather than focusing on inherent characteristics such as climate and physical distance to the coast and market areas, we analyze the economic geography of the region. Economic geography characteristics include two elements: *market access*, represented by the transport network linking a location to market centers; and *spatial externalities*, represented by the local presence of buyers and suppliers to facilitate inter-industry transfers, the local presence of firms in the same industry to facilitate intra-industry transfers, and the diversity of the local industrial base. Drawing on testable hypotheses from the New Economic Geography (NEG) literature, this analysis provides the micro-foundation for understanding whether a region's economic geography influences location decisions at the firm level. Only by first explaining these decisions, will it be possible to build a general framework for evaluating the overall spatial distribution of economic activity and employment.

2

Using plant or "factory" level data for 1998-99, from the Indian Annual Survey of Industries (ASI), we examine location choices in eight manufacturing industries.[1] These are (with National Industrial Classification [NIC] codes in parenthesis):

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)

These plant level data are supplemented by district and urban demographic and amenities data from the 1991 Census of India and detailed, geographically referenced information on the availability and quality of transport infrastructure linking urban areas (CMIE, 1998; ML Infomap, 1998). The Annual Survey of Industries (ASI) data allow us to identify each plant at the district level spatially and at the four digit SIC level sectorally.

This paper is organized in three parts. In Part I, we present the analytic framework and specify the econometric model to examine location decisions at the firm level.[2] In Part II, we discuss results from the econometric analysis. Part III briefly summarizes the contributions and

---

[1] By grouping firms into carefully defined sectors (rather than examining all manufacturing together), we can identify the differential impact of regional characteristics or geographic externalities across industries. For example, in comparison to Food Processing, which is closely linked to the traditional rural industrial base, industries such as Machinery, Metals, and Computers and Electronics are relatively footloose urban industries subject to considerable agglomeration economies.

[2] The ASI provides information on plants or factories, which are the units of production. These are roughly equivalent to the use of establishment level data. The industry survey does not allow us to identify enterprises to whom individual establishments may be linked.

3

implications of the findings. The principal conclusion is that only one economic geography variable—local economic diversity—has significant, consistent, and substantial cost-reduction effects. Hence, diversity matters.

I. MODEL OF INDUSTRY LOCATION

1.1 Analytic Framework

The analytic framework to examine location of manufacturing industry primarily draws on recent findings from the NEG literature. In the 'new economic geography' literature Krugman (1991a, 1991b) and Fujita et al. (1999) analytically model increasing returns, which stem from mostly pecuniary externalities.[3] They emphasized the importance of supplier and demand linkages and transportation costs. Firms prefer to produce each product in a single location given fixed production costs, and firms also prefer to locate their production facilities near large markets to minimize transportation costs.

Drawing upon Fujita and Thisse (1996) and Fujita (1989), we model firms to benefit from externalities arising from being co-located with other firms. If $a(x, y)$ is the benefit to a firm at x obtained from a firm at y, and $f(y)$ denotes the density of firms at each location $y \in X$ then,

$$A(x) \equiv \int_X a(x,y) f(y) dy \tag{1}$$

---

[3] NEG's approach bears a strong resemblance to Marshall (1890) and Weber (1929) in many ways. However, unlike its predecessors, new economic geographers place less emphasis on technology spillovers as a source of externalities than on labor pooling and specialized suppliers. Krugman (1991) argued that externalities from technology spillovers are difficult to measure, and therefore, cannot be modeled. Instead, he argued that under increasing returns to scale and imperfect competition, pecuniary externalities have clear welfare effects due to the variety of market size effects (i.e., each firm's monopoly power can affect the production function of other firms through buying and selling in the market) (Krugman, 1993). By focusing on pecuniary externalities (or rent spillovers) rather than technology spillovers, NEG tries to focus the general discussion on externalities.

Thus, A(*x*) represents the aggregate benefit accrued to a firm at *x* from the externalities created in location *X*. Assuming that production utilizes land (S$_f$) and labor (L$_f$) with rents of R(*x*) and W(*x*) respectively at *x*, a firm located at $x \in X$ would maximize profits subject to:

$$\Pi(x) = A(x) - R(x)S_f - W(x)L_f \qquad (2)$$

Note that, as an aggregate term, the density of firms at each location, *f(y)*, can represent regional economic attributes based on inter-firm relationships (in other words economic geography). Specifying types of such attributes unpacks the sources of spatial externalities, which have been often treated as a black box in neoclassical urban system models (Henderson 1974, 1977, 1988). First, a large geographic concentration of similar firms can provide scale economies in the production of shared inputs. Besides, firms that utilize similar technologies and face common issues are more likely to collaborate with one another to share information on a variety of issues from problem solving to the development of new production technologies. Second, the benefits from locating near own industry concentrations can be augmented by the presence of inter related industries. To a large extent, the work on inter-industry externalities have been motivated by research on industry clusters. Clusters can be defined as a geographically concentrated and interdependent network of firms linked through buyer-supplier chains or shared factors. The success of an industry cluster hinges on how well such local linkages among firms, education and research institutions, and business associations can be developed. The 'cluster' concept particularly emphasizes interfirm relations that reduce the cost of production by lowering transaction costs among firms (Porter 1990). Interrelated firms located in proximity can reduce their transportation cost for intermediate goods and can share valuable information on their products more easily. Therefore, for profit maximizing firms, the presence of a well-developed network of suppliers in a region is an important factor for their location decision. Lastly, economic diver-

sity of a region is another important source of spatial or location based externalities. Firms located in larger metro area are more likely to benefit not only from inter-industry technology spillovers but also from easier access to producer services such as legal services or banking.

Transport costs are also important in determining the location choice of firms. Krugman (1991b) shows that manufacturing firms tend to locate in regions with larger market demand to realize scale economies and minimize transportation cost. If transport costs are very high, then activity is dispersed. In the extreme case, under autarky, every location must have its own industry to meet final demand. On the other hand, if transport costs are negligible, firms may be randomly distributed as proximity to markets or suppliers will not matter. Agglomeration would occur at intermediate transport costs when the spatial mobility of labor is low (Fujita and Thisse 1996). We therefore expect a bell shaped (inverted U shaped) relationship between the extent of spatial concentration and transport costs.

To include transport costs in a firm's location decision, we modify equation (2) as:

$$\Pi(x) = A(x) - R(x)S_f - W(x)L_f - TC(x) \tag{3}$$

where $TC(x)$ represents the transport costs of the firm at location x. With a decline in transport costs, firms have an incentive to concentrate production in a few locations to reduce fixed costs. Transport costs can be reduced by locating in areas with good access to input and output markets. Thus, access to markets is a strong driver of agglomeration towards locations where transport costs are low enough that it is relatively cheap to supply markets. In addition to the pure benefits on minimizing transport costs, the availability of high quality infrastructure linking firms to urban market centers increases the probability of technology diffusion through interaction and knowledge spillovers among firms, and also increases the potential for input diversity (Lall et al., forthcoming). Analytical models of monopolistic competition generally show that

6

activities with increasing returns at the plant level are pulled *disproportionately* towards locations with good market access.

The analytic framework in this section highlights the importance of economic geography in influencing location and agglomeration at the firm level. Insights from NEG and regional science models suggest that own and inter-related industry concentrations, availability of reliable infrastructure to reduce transport costs and enhance market access, regional amenities and economic diversity are important for reducing costs, thereby influencing location and agglomeration of industry. In Section 1.2, we describe the economic geography variables that are used in this analysis. The econometric specification to evaluate the importance of these variables is described in Section 1.3. The empirical strategy is to estimate a cost function to see how costs (thereby profits) are affected by the economic geography of the region where the firm is located. If specific factors related to the local economic geography have cost reducing impacts, then firms are likely to choose regions with disproportionately higher levels of these factors.

1.2 Economic Geography Variables

Own industry concentration:

The co-location of firms in the same industry (localization economies) generates externalities that enhance productivity of all firms in that industry. These benefits include sharing of sector specific inputs, skilled labor, and knowledge, intra-industry linkages, and opportunities for efficient subcontracting. Firms that share specialized inputs and production technologies are more likely to cooperate in a variety of ways. In many industries, it is common for competitors in the market to launch joint projects for new product and process development. Further, a dis-

proportionately high concentration of firms within the same industry increases possibilities for collective action to lobby regulators or bid-prices of intermediate products.

There is considerable theorizing on localization economies in the works of Marshall (1890), Arrow (1962), and Romer (1986). They argue that cost-saving externalities are maximized when a local industry is specialized (often called MAR externalities), and their models predict that externalities predominantly occur within the same industry. Therefore, if an industry is subject to MAR externalities, firms are likely to locate in a few cities where producers of that industry are already clustered. Examples of highly localized industries are ubiquitous. Semiconductor and software in Silicon Valley and automobile in Detroit are classic cases in point. Later, Porter (1990) also emphasized the importance of dynamic externalities created in specialized and geographically concentrated industries.

There is an extensive empirical literature supporting the positive effects of localization economies (Henderson 1988, and Ciccone and Hall 1995). In a recent study of Korean industry, Henderson et al. (1999) estimate scale economies using city level industry data for 1983, 1989, and 1991-93, and find localization economies of about 6 to 8 percent. However, while industry concentration provides many benefits, some of these may be offset by costs from enhanced competition between firms for labor and land causing wages and rents to rise, as well as higher transport costs due to congestion. Therefore, the net benefits of own industry concentration may be marginal for sectors with low skilled labor and standardized technologies.

There are several ways of measuring localization economies. These include own industry employment in the region, own industry establishments in the region, or an index of concentration, which reflects disproportionately high concentration of the industry in the region in comparison to the nation. We use own industry employment in the district to measure localization

economies. This measure is consistent with the type of benefit spillovers specified in equation

(1), where localization economies come from the absolute volume of other activity in the district.

Own industry employment is calculated from employment statistics provided in the 1998-99

sampling frame of the ASI, which provides employment data on the universe of registered indus-

trial establishments in India. The sample data used for the cost function estimation are drawn

from this sampling frame.

Inter-Industry Linkages

In addition to intra-industry externality effects, we also include a measure to evaluate the

importance of inter-industry linkages in explaining firm level profitability, and thereby location

decisions. The importance of inter-industry linkages as a major agglomerative force was first in-

troduced by Marshall (1890, 1919). Venables (1996) recently demonstrated that agglomeration

could occur through the combination of firm location decisions and buyer-supplier linkages even

without high factor mobility. The presence of local suppliers can reduce transaction costs and

therefore increase productivity. Inter-industry linkages can also serve as a channel for vital in-

formation transfers. Firms that are linked through stable buyer-supplier chains often exchange

ideas on how to improve the quality of their products or on how to save production costs. It is

such on-going interactions that make the dynamics of inter-industry externalities so vibrant.

Therefore, if the performance of an industry is highly dependent upon the supply of high-quality

intermediate goods (e.g., automobile manufacturing), firms are likely to locate in regions with a

strong presence of local suppliers. The presence of local supplier linkages makes buyer industries

more efficient and reinforces the localization process.

9

There are several approaches for defining inter-industry linkages: input-output based, labor skill based, and technology flow based. Although these approaches represent different aspects of industry linkages and the structure of a regional economy, the most common approach is to use the national level input-output accounts as templates for identifying strengths and weaknesses in regional buyer-supplier linkages (Feser and Bergman 2000). The strong presence or lack of nationally identified buyer-supplier linkages at the local level can be a good indicator of the probability that a firm is located in that region.

To evaluate the strength of buyer-supplier linkages for each industry, we use the summation of regional industry employment weighted by the industry's input-output coefficient column vector from the national input-output account:

$$L_{ir} = \sum_{i=1}^{n} \omega_i e_{ir} \tag{4}$$

where L_{ir} is the strength of the buyer supplier linkage, ω_i is industry i's national input-output coefficient column vector and e_{ir} is total employment for industry i in district r. This is similar to the measure used in Koo (2002) to define the strength of buyer-supplier chains. The measure examines local level inter-industry linkages based on the national input-output accounts. The national I-O coefficient column vectors describe intermediate goods requirements for each industry (i.e., inter-industry linkages). Assuming that local industries follow the national average in terms of their purchasing patterns of intermediate goods, national level linkages can be imposed to the local level industry structure for examining whether region r has a right mix of supplier industries for industry-i. By multiplying the national input-output coefficient column vector for industry i and the employment size of each sector in region r (a district is used as a geographical unit for buyer-supplier linkage analysis), simple local employment numbers can be weighted based on what industry i purchases nationally.

10

Indeed, the importance of local linkages is determined by the size of its industrial base

(e.g., employment in each industry) and the extent to which local industries can provide interme-

diate goods for local firms (from the IO coefficient vector). In this case, our measure takes two

important aspects of buyer-supplier linkages into account--fit and size. While computing the in-

dicator, we noticed that the industry categories in the NIC system and in IO accounts do not have

an exact match. Therefore, we first developed a concordance table between them before multi-

plying w_i and e_{ir}. Data on input output transactions are from the Input Output Transactions Table

1993-94, Ministry of Statistics and Programme Implementation.

Economic Diversity

In addition to buyer-supplier linkages, there are other sources of inter-industry external-

ities. Prominent among these is the classic Chinitz-Jacobs diversity. The diversity measure pro-

vides a summary measure of urbanization economies, which accrue across industry sectors and

provide benefits to all firms in the agglomeration. Chinitz (1961) and Jacobs (1969) proposed

that important knowledge transfers primarily occur across industries and the diversity of local

industry mix is important for these externality benefits. They argue that cities are breeding

grounds for new ideas and innovations due to the diversity of knowledge sources concentrated

and shared in cities. The diversity of cities facilitates innovative experiments with an array of

processes, and therefore new products are more likely to be developed in diversified cities.

Therefore, industries with Jacobs type externalities tend to cluster in more diverse and larger

metro areas. (Recently, Duranton and Puga (1999) designed a model providing the micro-

foundations of a Jacobs-type model.) The benefits of locating in a large diverse area go beyond

the technology spillovers argument. Firms in large cities have relatively better access to business

services, such as banking, advertising, and legal services. Particularly important in the diversity argument is the heterogeneity of economic activity. On the consumption side, increasing the range of local goods that are available enhances the utility level of consumers. At the same time, on the production side, the output variety in the local economy can affect the level of output (Abdel-Rehman 1988, Fujita 1988, Rivera Batiz 1988). That is, urban diversity can yield external scale economies through the variety of consumer and producer goods. Recent empirical studies by Bostic et al. (1997) and Garcia-Mila and McGuire (1993) show that diversity in economic activity has considerable bearing on the levels of regional economic growth. The later type of benefit is particularly important in developing countries, where most manufacturing industries are based on low skills and low wages but abundant local labor forces.

In this study, we use the well-known Herfindahl measure to examine the degree of economic diversity in each district. The Herfindahl index of a region r (H_r) is the sum of squares of employment shares of all industries in region r:

$$H_r = \Sigma_j (\frac{E_{jr}}{E_r})^2 \qquad (5)$$

Unlike measures of specialization, which focus on one industry, the diversity index considers the industry mix of the entire regional economy. The largest value for H_r is one when the entire regional economy is dominated by a single industry. Thus a higher value signifies lower level of economic diversity. Therefore, for more intuitive interpretation of the measure, for the diversity index in our model, H_r is subtracted from unity. Therefore, $DV_r=1-H_r$. A higher value of DV_r signifies that the regional economy is relatively more diversified.

The results from empirical studies on the relative importance of specialization and diversity are mixed. Glaeser *et al.* (1992) find evidence only in favor of diversity. On the other hand, Miracky (1995) finds little evidence to support the diversity argument. Henderson *et al.* (1995)

show that the relative importance depends on the choice of industry. They find evidence of specialization externalities in mature capital goods industries and of diversity externalities in new high-tech industries. These findings are consistent with the product cycle theory (Vernon 1966) which predicts that new industries tend to prosper in large and diverse urban area, but with maturity, their production facilities move to smaller and more specialized cities.

Market Access (MA)

In principle, improved access to consumer markets (including inter-industry buyers and suppliers) will increase the demand for a firm's products, thereby providing the incentive to increase scale and invest in cost reducing technologies. The distance from and the size and density of market centers in the vicinity of the firm determine access to markets. The classic gravity model, which is commonly used in the analysis of trade between regions and countries (Evennet and Keller 2002), states that the interaction between two places is proportional to the size of the two places as measured by population, employment or some other index of social or economic activity, and inversely proportional to some measure of separation such as distance. Following Hansen (1959):

$$I_i^c = \sum_j \frac{S_j}{d_{ij}^b}$$
(6)

where I_i^c is the 'classical' accessibility indicator estimated for location i, S_j is a size indicator at destination j (for example, population, purchasing power or employment), d_{ij} is a measure of distance (or more generally, *friction*) between origin i and destination j, and b describes how increasing distance reduces the expected level of interaction. Empirical research suggests that simple inverse distance weighting describes a more rapid decline of interaction with increasing dis-

tance than is often observed in the real world (Weibull, 1976). The most commonly used modified form is a negative exponential model such as:

$$I_i^{ne} = \sum_j S_j \cdot e^{\left(-d_{ij}^b / 2a^2\right)}$$ (7)

where I_i^{ne} is the potential accessibility indicator for location i based on the negative exponential distance decay function, most other parameters are defined as before, and the parameter a is the distance to the point of inflection of the negative exponential function.

There are several options for developing accessibility indicators depending on the choice of distance variables used in the computation. These include: (a) indicators based on Euclidean distance; (b) indicators incorporating topography; (c) indicators incorporating the availability of transport networks; and (d) indicators incorporating the quality of transport networks. In this analysis, we use network distance as the basis of the inverse weighting parameter and to incorporate information on the quality of different transportation links. Feasible travel speed and thus travel times will vary depending on each type of network link. A place located near a national highway will be more accessible than one on a rural, secondary road. The choice of the friction parameter of the access measure will therefore strongly influence the shape of the catchment area for a given point—i.e., the area that can be reached within a given travel time. This, in turn, determines the size of potential market demand as measured by the population within the catchment area.

We use the accessibility index developed in Lall et. al (forthcoming) as the market access indicator in this analysis. Their accessibility index describes market access using information on the Indian road network system and the location and population of urban centers (ML Infomap 1998). The urban centers database includes latitude and longitude coordinates and 1991 population for 3,752 cities with a total population of about 217 million. The digital transport network

14

data set includes an estimated 400,000 km of roads. Each road segment is categorized into four classes according to road quality: national highways (about 30,000 km or 7.7 percent of total roads), state highways (90,000 / 22.5 percent), secondary connector roads (120,000 / 29.8 percent), and other roads (160,000 / 39.9 percent). The complete digital representation of India's transport network that is used for the accessibility index thus consists of a set of urban centers represented as nodes connected by lines that correspond to roads of different quality. Rather than distance, the weighting parameter used in the accessibility computation is an estimate of travel time. This makes it possible to incorporate road quality information by assigning different travel speed estimates to different types of roads. Based on information available in the Indian Infrastructure Handbook (CMIE, 1998) varying travel speeds ranging from 25 to 50 km/hour were used, depending on the type of road. The algorithm used to compute the accessibility measure is based on the familiar Dijkstra algorithm. We use this to compute the network travel time to urban centers for each of more than 100,000 points distributed across India. As the exact geographic location of each firm is not publicly available, we summarized the accessibility for each district by averaging the individual values for all points that fall into the district. The negative exponential function in Equation (7) is chosen as the most suitable functional form for the decay of interaction with increasing travel time.

1.3 Econometric Specification

In this section, we present the econometric specification to test the effects of economic geography factors in explaining the location of economic activity. Our basic premise is that firms will locate in a particular location if profits exceed some critical level demanded by entrepreneurs. We estimate a cost function with a mix of micro level factory data and economic geography

15

variables, which may influence the cost structure of a production unit. After developing the estimation methodology, we also provide a short description of the data sources.

A traditional cost function for a firm i is (subscript i is dropped for simplicity):

$$C = f(Y, w) \qquad (8)$$

where C is the total cost of production for firm i, Y is its total output, w is an n-dimensional vector of input prices. However, the economic geography, or the characteristics of the region where the firm is located, is also an important factor affecting the firm's cost structure. The production cost of a firm is determined not only by its output and the value of its inputs, but also by ease of access to markets via reliable transportation networks, availability of a diverse input mix, and technological externalities from similar firms located in the region. Such location-based advantages have clear implications for a firm's location decision as they create cost-saving externalities. We modify the basic cost function to include the influence of location-based externalities:

$$C_r = f(Y, w_r, A_r) \qquad (9)$$

where C_r is the total cost of a firm in region r, w_r is an input price vector for the firm in district r, and A is a m-dimensional vector of spatial externalities (i.e., economic geography or agglomeration variables such as access to markets, buyer supplier networks, own industry concentration) at location r.

The model has four conventional inputs: capital, labor, energy, and materials. Therefore, the total cost is the sum of the costs for all four inputs. With respect to agglomeration economies, it is assumed that there are four sources of agglomeration economies at the district level (described in the previous section) such that $A=\{A_1, A_2, A_3, A_4\}$, where A_1 is the market access measure, A_2 is the concentration of own industry employment, A_3 is the strength of buyer-supplier linkages, and A_4 is the relative diversity in the region.

16

Shephard's lemma produces the optimal cost-minimizing factor demand function for input j corresponding to input prices as follows:

$$X_{jr} = \frac{\partial C_r}{\partial w_{jr}}(Y, w_r, A_r) \qquad j = 1,2,3,4,....,n \qquad (10)$$

where X_{jr} is the factor demand for j^{th} input of a firm in district r. It is clear that the firm's factor demand is determined by its output, factor prices, and location externalities. Therefore, the production equilibrium is defined by a series of equations derived from equation (9) and (10).

The empirical implementation of the above model is based on a translog functional form, which is a second-order approximation of any general cost function. Since there are four conventional inputs and four location externalities (agglomeration) variables, a translog cost function can be written as:

$$\ln C = \alpha_0 + \alpha_y \ln Y + \sum_j \alpha_j \ln w_j + \sum_l \alpha_l \ln A_l + 1/2\beta_{yy}(\ln Y)^2 + 1/2\sum_j \sum_k \beta_{jk} \ln w_j \ln w_k +$$
$$\sum_j \beta_{jy} \ln Y \ln w_j + 1/2\sum_l \sum_q \gamma_{lq} \ln A_l A_q + \sum\sum \gamma_{jl} \ln w_j A_l + \sum_l \gamma_{ly} \ln Y \ln A_l \qquad {}^{(j \neq k;}$$

$$l \neq q; j,k=1,2,3,4 ; l,q=1,2,3,4) \qquad (11)$$

In addition, from equation (10), the cost share of input factor j can be written as

$$S_j = \alpha_j + \sum_k \alpha_{jk} \ln w_k + \beta_{jy} \ln Y + \sum_l \gamma_{jl} A_l \qquad (k=1,2,3,4;l=1,2,3) \qquad (12)$$

Notice that the cost share equations of all factor inputs satisfy the adding up criterion, $\sum_j S_j = 1$. The 'adding up criterion' has important implications for model estimation. The system of cost share equations satisfies the 'adding up criteria' if

$$\sum_j \alpha_j = 1; \sum_k \beta_{jk} = \sum_j \beta_{jk} = 0; \sum_j \beta_{jy} = 0; \sum_l \gamma_{jl} = \sum_j \gamma_{jl} = 0 \qquad (13)$$

thereby, reducing the number of free parameters to be estimated.

The translog cost function can be directly estimated from equation (11). However, a joint estimation of equation (11) and (12) with restriction (13) significantly improves the efficiency of

the model. The final model estimated includes two additional dummy variables that identify locational characteristics that may not be captured by agglomeration variables. Locations are categorized as rural, nonmetro urban (D_1), and metro urban (D_2), and rural location is used as a reference category.

The impact of the economic geography factors on the cost structure (or profitability) of the firm can be evaluated by deriving the elasticity of costs with respect to the economic geography variables. From equation (11) the cost elasticities are:

$$\frac{\partial C}{\partial A_l} = \alpha_l + \Sigma_j \gamma_{jl} \ln w_j + \Sigma_q \gamma_{lq} \ln A_q + \gamma_{ly} \ln Y \tag{14}$$

In addition to direct impacts on the cost structure, these location specific externalities also influence factor demand. The impact of these variables on input demand can be derived from the cost share equations. Note that the cost share for input j, S_j, can be written as $w_j v_j / C$, where w_j is factor price of input j, v_j is the quantity demanded of input j, and C is total cost. That is,

$$v_j = \frac{C}{w_j} S_j \text{ and } \ln v_j = \ln C + \ln S_j - \ln w_j \tag{15}$$

Therefore, the elasticities of input demands with respect to agglomeration factors A_l is

$$\frac{\partial \ln v_j}{\partial \ln A_l} = \frac{\partial C}{\partial A_l} + \frac{\gamma_{jl}}{A_l} \tag{16}$$

18

1.4 Data

Data Sources

We use plant level data for 1998-99 from the Annual Survey of Industries (ASI), conducted by

the Central Statistical Organization of the Government of India.[4] The "factory" or plant is the

unit of observation in the survey and data are based on returns provided by factories.[5] Data on

various firm level production parameters such as output, sales, value added, labor cost, employ-

ees, capital, materials and energy are used in the analysis. In summary, factory level output is

defined as the ex-factory value of products manufactured during the accounting year for sale.

Capital is often measured by perpetual inventory techniques. However, this requires tracking the

sample plant over time. This is a major task for micro-level research due to changes in sampling

design and incomplete tracking of factories over time. Instead, in our study (and in the ASI data-

set) capital is defined as the gross value of plant and machinery. It includes not only the book

value of installed plant and machinery, but also the approximate value of rented-in plant and ma-

chinery. Doms (1992) demonstrates that defining capital as a gross stock is a reasonable ap-

proximation for capital. Labor is defined as the total number of employee person-days worked

and paid for by the factory during the accounting year.

The factory or plant level data from the Indian ASI allows us to compute input costs.

With respect to input costs and input prices, capital cost is defined as the sum of rent paid for

land, building, plant, and machinery, repair and maintenance cost for fixed capital, and interest

on capital. Labor cost is calculated as the total wage paid for employees. Energy cost is the sum

---

[4] The ASI covers factories registered under sections 2m(i) and 2m(ii) of the Factories Act 1948, employing 10 or more workers and using power, and those employing 20 or more workers but not using power on any day of the preceding 12 months.

[5] Goldar (1997) notes that factories are classified into industries according to their principal products. In some cases this causes reclassification of factories from one class to another in successive surveys, making inter-temporal comparisons difficult.

of electricity (both generated and purchased), petrol, diesel, oil, and coal consumed. The value

of self-generated electricity is calculated from the average price that a firm pays to purchase

electricity. Material cost is the total aggregate purchase value for domestic and foreign interme-

diate inputs. We define the price of capital as the ratio of total rent to the net fixed capital. The

price of labor is calculated by dividing total wage by the number of employees. Energy and ma-

terial prices are defined as weighted expenditure per unit output. Output value is weighted by

factor cost shares.

Data quality has been examined by cross referencing with standard growth accounting

principles as well as by reviewing comments from other researchers who have used these data.

The geographic attributes allow us to identify each firm at the district level.

Data Imputation:

The 1999 ASI data used for estimation have a significant number of incomplete cases.

Many firms did not report their capital, output, depreciation, and other related input price infor-

mation. Even when there are reported values, some of them are not consistent (e.g., 0 capital

when capital depreciation is reported positive). Missing or inconsistent data can be a serious

problem when such data points are not completely random.

To take into account the limitations arising from the less than perfect ASI data, we first

adopted the following set of rules to clean the data, and then imputed missing values in the

cleaned data using SAS MI procedure. First, cases that are missing too much vital information

(e.g., input, output, capital, and employment) are deleted (only 78 cases were deleted from this

step). Second, when the value for plant and machinery depreciation is positive and the size of

employment is greater than 10, but the closing value of capital is reported 0, capital is converted

to missing. Lastly, when capital is missing, but its depreciation value is 0, depreciation is converted to missing because it is likely that newly imputed values for capital will be positive which implies positive depreciation of capital.

The easiest and probably the most frequently used methods to handle missing data points are casewise data deletion and mean substitution. If a case has any missing values, the entire record can be deleted or missing points can be substituted by mean values. However, Roth (1994) compared different approaches often used in empirical research and concluded casewise data deletion and mean substitution are inferior to maximum likelihood based methods such as multiple imputation.

To resolve the issue of missing data, we introduce a multiple imputation technique developed by Rubin (1978, 1987) and others. The multiple imputations usually generate five to ten complete data sets by filling in gaps in existing data with proper raw values. Raw values are drawn from their predicted distribution based on the observed ones. Then each complete data set can be analyzed by common statistical methods (e.g., regression). After conducting identical analysis multiple times, the results drawing upon imputed data sets are combined into one summary set of parameters.

We generated five complete data sets and used mean values to impute missing cases. The imputed values were evaluated again to check their consistency. When imputed values were unreasonably small or large, we converted them back to missing and imputed again. The imputation procedure was repeated three times.

These plant level data are supplemented by district and metropolitan area level demographic and amenities data from the 1991 Census of India and detailed information on the availability and quality of transport infrastructure linking urban areas. The plant level data have been

combined with district level indicators such as concentration of industry in the district, urban population density, and potential access to urban markets.

1.5 Summary information on spatial variation

Before moving on to discussing the results from the empirical analysis, we provide a general overview on the concentration and basic characteristics of firms in the study sectors. We first divide the economic landscape to comprise of non urban areas, urban areas, and large metropolitan areas. The metropolitan areas include the following cities and their urban agglomerations—Delhi, Mumbai, Kolkata, Chennai, Bangalore, and Ahmedabad. Using the sample data from the ASI for 1998-99, we see that average wages across industries are the highest in metropolitan areas (see Table 1). In comparison to a nationwide average annual of Rs. 60,000 per employee, labor remuneration is Rs. 74,000 for metropolitan areas, Rs. 54,000 for other urban areas and Rs. 50,000 for non urban areas. Among various industries, annual wages are the highest in electronics and computers (Rs. 101,000 per employee) and lowest in the leather industry (Rs. 41,000 per employee). Even within sectors (not shown in this table), wages tend to be higher as we move up the urban scale.

Productivity indicators such as output per employee and value added per employee show interesting trends. While per employee output is relatively quite high in several industries, the value added figures show quite a different situation. For example, per employee output in computing and electronics is Rs. 344,000 but value added per employee is only Rs. 65,000. Similarly, the numbers for output and value added per employee are Rs. 376,000 and Rs. 79,000 for

Chemicals and Rs. 314,000 and Rs. 204,000 for printing and publishing. This suggests that these industry sectors are not very efficient in transforming inputs into higher value outputs.

Next, we use the Ellison-Glaeser (1997) index of concentration to see if industrial activity within sectors is clustered across locations. Their concentration index can be defined as:

$$r = \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1 - \sum_{i}^{M} x_i)H}{(1 - \sum_{i}^{M} x_i)(1 - H)} \tag{17}$$

where r is the extent to which an industry is geographically concentrated, s_i is the region i's share of the study industry, x_i is the regional share of the total employment, and H is the Herfindahl industry plant size distribution index, $H = \sum_{j=1}^{N} z_j^2$. The EG index is explicitly derived from the micro foundations of a firm's location choice. It takes on a value close to zero when the distribution of plant location is completely random (as opposed to a uniform distribution). Therefore, a non zero value implies agglomeration or clustering above and beyond what we would observe if the firm's location decisions are random (in general, an industry is highly concentrated if $r \geq$ 0.05, moderately concentrated if r is between 0.02 and 0.05, and not concentrated if r <0.02). The index is designed to allow comparisons across industries, countries, and over time. Therefore, in principle it is possible to compare the concentration of industries in the U.S. and Mexico or that of high-tech and low-tech industries.

The EG concentration index in Table 2 is computed at the state level, using data from the sampling frame of the ASI. Therefore, the employment summaries in Table 2 reflect the universe of all employment in registered establishments. We find evidence of high spatial concentration for the Leather and Metals sectors, and moderate concentration in Food Products, Textiles, Mechanical Machinery and Computing and Electronics. Firms in the Paper Products and Chemicals sectors do not exhibit patterns of spatial concentration.

II. RESULTS FROM ECONOMETRIC ANALYSIS

The empirical analysis is conducted by jointly estimating equations (11) and (12) as a system,

using an iterative seemingly unrelated regression (ITSUR) procedure. The underlying system is

nonlinear, and is primarily derived from the structure of the input demands, as represented in

equation (11). The ITSUR procedure estimates the parameters of the system, accounting for het-

eroscedasticity, and contemporaneous correlation in the errors across equations. As the cost

shares sum to unity, *n-1* share equations are estimated (where *n* is the number of production fac-

tors). The ITSUR estimates are asymptotically equivalent to maximum likelihood estimates and

are invariant to the omitted share equation (Greene, 1993). All estimations were carried out with

the MODEL procedure of the SAS system.

It is quite likely that due to heterogeneity in technology use, production efficiency, and

managerial capacity among firms of different sizes, it may be limiting to group all firms in the

same estimation process. Further, the benefits of location specific characteristics may be accrued

more by smaller firms, who are relatively more dependent on access to buyers and suppliers,

availability of ancillary services, inter firm non-technological externalities, and high quality in-

frastructure. In contrast, larger firms may be in a better position to internalize production of vari-

ous intermediate goods, self-provide infrastructure, and stock higher inventories. As a result,

they are relatively less dependent on location based amenities and characteristics. To make al-

lowances for this heterogeneity, and test if in fact there are differences in production costs and

the impact of economic geography across firms of different sizes, we classify firms into three

categories: small, medium, and large. Small firms are defined as those with less than 50 employ-

ees, medium sized are between 50 and 99 employees and large firms have 100 or more employees. The number of firms by size category is reported in Table 3.

Summary results for the estimated cost functions are reported in Tables 4 and 5. Table 4 provides results for the conventional inputs (capital, labor, energy and materials) and Table 5 provides estimates for the economic geography variables.[6] We present these separately as the economic geography variables are external effects, not directly included in the firm's cost structure. In these tables, we provide results for the industry in general, followed by specific parameter estimates for small, medium, and large firms. From Table 4 it is quite clear that increase in factor prices translates into higher overall costs at the firm level.

Table 5 summarizes the impact of the economic geography factors on the cost structure (or profitability) at the level of the firm. The estimates in Table 5 are the cost elasticities of these variables, as defined in equation (14). There are four sets of location/ economic geography variables in the analysis: (a) access to markets (Access), (b) own industry concentration (Emp), (c) buyer supplier or input output linkages (IO link), and (d) local economic diversity (Diversity). The results for each industry sector are provided in four parts. The first column has industry-wide cost elasticities. These are followed by estimates for small, medium, and large firms respectively. As we see from the results, sorting by firm size helps us identify particular types of firms, which are likely to benefit more from location based characteristics. In general, the cost elasiticities show that there is considerable heterogeneity in the impact of location characteristics on costs incurred at the firm level. This heterogeneity is not limited to the overall effects across in-

---

[6] There are some cells in Tables 4 and 5 with no values. We do not report the estimated parameters in these cases as the number of observations (see Table 3) is too few to allow any meaningful interpretation of the results – especially when the model estimates around 50 parameters. As a rule of thumb, we do not report results for estimations with less than 200 observations (firms).

dustries, but also includes differences across firms of different sizes and by sources of agglomeration economies.

We start by describing the impact of access to markets. Market access, measured by transport network quality and urban population, measures effective demand for a firm's products and the ease with which it can reach buyers and suppliers. Locating in a region with good access to markets is likely to reduce the cost of intermediate inputs as well as increase demand for the firm's products. This will provide the entrepreneur with incentives to increase scale of production and also invest in cost reducing technologies (Lall et. al forthcoming). The industry-wide results for market access suggest that that the net cost reducing impact of market access is not significant in most industry sectors. The estimated cost elasticities are negative and statistically significant for two industry sectors – metals and mechanical machinery – the elasticity values are insignificant for other sectors. For example, in Mechanical Machinery, the coefficient of –0.046 means that a 10 % improvement in market access will be associated with an approximately 0.5% reduction in overall costs at the firm level. We get a counter intuitive result for the leather industry, where the cost elasticity is positive and significant.

For small firms however, the estimated elasticities are generally negative, indicating benefits from improved market access. However, the estimates are statistically significant at the 5% level for only two industry sectors – chemicals and metals. We also find a positive and significant estimate for the Textiles industry, suggesting that there are costs associated with higher market access. Most of the estimates for medium and large industries are not statistically significant.

Following market access, we discuss results for own industry concentration, which is measured as the sum of employment in the particular industry in the region. As in the case of

market access, the reported estimates are elasticities, derived following the specification in equation (7). The industry-wide estimates suggest that there are no net benefits of being located near own industry concentrations. All the estimated elasticities are positive, which suggests that costs increase if firms locate in regions with high concentrations of the same industry. These coefficients are statistically significant at the 1% level for four sectors and significant at the 5% for one industry sector. To examine if industry wide results are artifacts of aggregation, it is useful to look at the results by firm size. We find that even when disaggregated by firm size, own industry concentration systematically provides either no net benefits and, in some instances, actually increases costs at the firm level.

The findings for input output linkages (IO link) show that for most industry sectors, proximity to buyers and suppliers potentially reduces costs at the firm level. While the estimated elasticities are negative for six sectors, it is only statistically significant at the 5% level for the metals industry. The coefficient of −0.01 means that a 10 % increase in the strength of buyer supplier linkages is associated with firm level cost reductions of 0.1%. Or in other words, doubling the strength of buyer supplier linkages is associated with a 1% reduction in firm level production costs. When we look at the elasticities for small firms, we find that the estimates are insignificant in most cases. For medium size firms, the elasticity is negative and significant for the metals sector. The coefficient of 0.17 means that a doubling of IO linkages is associated with a 17% reduction in firm level costs. This effect is considerably stronger than the other estimates, where the cost elasticities rarely exceed 5%. For large firms, we find that costs increase for food and beverages and for electrical/electronics, when firms are located in regions with relatively higher buyer supplier linkages.

The estimates for local economic diversity indicate that there are considerable cost reducing benefits from being located in a diverse region. The industry wide estimates are negative for all sectors, and significant at the 1% level for the Food and Beverages and Textiles sectors. The coefficient of -0.10 for Textiles means that doubling of the region's economic diversity will reduce firm level costs by 10%. The results are even stronger for small firms. The estimated elasticities are negative for all industry sectors, and statistically significant for five sectors. What is really striking is the magnitude of these effects. For example, the estimated cost elasticity for electrical/electronics is 83% and for chemicals it is 46%. These estimates clearly suggest that there are considerable benefits of being located in a diverse economic region. The results for medium and larger firms however, do not show similar benefits for location in diverse economic regions. The cost reducing effects of being located in a diverse region are greater for small firms because they can rely on location based externalities to a larger extent than medium and big firms. The benefits come from better opportunities for subcontracting, access to a general pool of skilled labor, and access to business services, such as banking, advertising, and legal services. In addition to these pecuniary externalities, there are potential technological externalities from knowledge transfer across industries. Larger firms, being more vertically integrated and with higher fixed costs, are not likely to benefit from these externalities.[7]

In general, we find that the regional economic geography has a reasonable degree of impact on the cost structure of firms. The sources and the magnitudes of these impacts vary considerably across industry sectors. The only major source of benefits that are likely to influence location choice at the margin is the location's economic diversity. This is further likely to be the case

---

[7] While the estimated elasticity for large electrical/electronics firms is 235%, it is likely that this result is a statistical artifact, and driven by some outliers.

for small firms. The magnitude of the other effects are so small (elasticity values less than 5%),

that they are unlikely to influence firm location choices.

<Insert Table 6 here>

Results showing the effects of the economic geography factors on demand for traditional

inputs are presented in Table 6. The estimated values are elasticities of substitution for input

demands with respect to agglomeration factors, based on the specification in equation (16).

Briefly, the following points may be highlighted:

(a) In general, economic geography factors have negligible substitution effects on capital.

In most cases, coefficient estimates are negative, which implies cost-saving effects of economic

geography factors, but statistically insignificant. The only exceptions are the chemicals sector

and the electrical/electronics sector where higher market access, I-O links, and diversity lower

the capital requirements, especially for small firms. For instance, doubling market access or IO

linkages will reduce overall capital demand by 24.8 percent and 17.7 percent respectively. The

fact that capital substitution effects are negligible is not surprising because we cannot expect

capital intensity or the cost of borrowing to vary over space.

(b) Labor requirements are consistently lowered with higher measures for the economic

geography variables (with the general exception of diversity) for the textile, printing/publishing,

chemicals, machinery, and electrical/electronics sectors. These effects are most consistent for

small firms, though the largest substitution effects are for large firms. Indeed, small firms are

more likely to benefit from economic geography factors by tapping into external economies scale

that such factors bring about. We believe that these effects are related to access to skilled labor;

skilled/productive labor is likely to be available in areas with better access, high own industry

concentration, diversity, etc. Hence it is possible to use smaller work forces in places with superior economic geography.

(c) Energy requirements, on the contrary, are increased with higher values for the economic geography variables. The coefficients are consistently significant in the textiles and machinery sectors, and generally significant in the food and metals sectors. This effect is probably related to the Byzantine energy pricing methods used by Indian state electricity boards. In most cases the cross-subsidy systems punish urban industrial consumers to reward agricultural and residential consumers. As a result energy costs are higher in urban/metropolitan areas even if energy requirements remain the same.

(d) The patterns of substitution for materials is inconsistent between sectors and firm sizes. The only consistently significant substitution effects are in the textiles sector, but even there we can some variation (different signs) for different methods of firm aggregation. It is not possible to find general explanations for what appears to be a random pattern.

Although the results only partially show that economic geography factors affect traditional input factor demands in a consistent way, which does not contradict previous findings. The elasticities of substitution of externality variables with respect to input demands are often inconsistent, especially for materials. For instance, Bernstein (1988) and Bernstein and Nadiri (1988)'s studies on R&D spillovers showed that the elasticities of substitution of R&D spillovers with respect to traditional inputs do not have consistent patterns among different industries.

III. CONCLUDING COMMENTS

In conclusion, we would like to highlight three points. First, the analytic strategy and empirical specifications used here are original, comprehensive, and generalizable. Though our work is mo-

tivated by development issues, and the findings contribute to the literatures on urban, regional, and industrial development, the methodology developed here is not limited to the analysis of developing countries only. This strategy can be applied to most firm level examinations of location decisions in any country. In this regard, this study is a significant advance in the spatial analysis of industrialization, and specially the large and growing field focusing on externalities, clustering, and increasing returns.

Second, the principal finding—that industrial diversity (that is, the local presence of a mix of industries) provides significant cost savings for individual firms, and is the only economic geographic variable to do so—raises serious questions about the validity of much theorizing on localization economies. Our analysis shows that this cost saving is the most significant factor for firms of all sizes and in all sectors of manufacturing industry. Other spatial factors that, in theory, have some productivity enhancing effects or cost benefits (such as local presence of own industry, local access to buyers and suppliers) are found to have little or no influence on profitability. In other words, localized external economies have no discernible cost benefits. Rather, generalized urbanization economies (manifested in local economic diversity) provide the agglomeration externalities that lead to industrial clustering in metropolitan and other urban areas.

Third, the policy implications of the findings are quite significant. Consider only the spatial policy issues. The findings on the traditional production inputs, especially those pertaining to energy costs, are important, but deserve a separate and detailed treatment. The validity of developing "specialized clusters" in remote areas, as instruments to promote regional development in lagging or backward regions, must be questioned. Such approaches have been implemented with limited success historically (witness the rise and fall of the "growth pole" concept), but have seen resurgence with the "Porter style" competitive advantage analysis. In contrast, policies that

31

encourage the creation and growth of mixed industrial districts are likely to be more successful than single industry concentrations. However, this is easier said than done, especially in remote or lagging regions. If location-related cost advantages are not related to market access (whereby dispersed infrastructure investments, particularly in transportation, do not favor lagging regions), or localization economies, it is difficult to see how manufacturing industry can become the engine of growth in lagging regions.

References

Abdel-Rahman, H. (1988). Product differentiation, monopolistic competition, and city size. *Regional Science and Urban Economics, 18*, 69-86.

Bostic, R. (1997). Urban productivity and factor growth in the late 19th century. *Journal of Urban Economics, 4*, 38-55.

Bernstein, J. (1988). Costs of production, intra- and interindustry R&D spillovers: *Canadian evidence. Canadian Journal of Economics*, 21, 3240-347

Bernstein, J. and Nadiri, M. I. (1988). Interindustry R&D spillovers, rates of return, and production in high-tech industries. *American Economic Review*, 78, 429-434

Chinitz, B. (1961). Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review, 51*, 279-289.

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

CMIE. (1998). *Infrastructure*. Bombay: Economic Intelligence Service, Center for Monitoring Indian Economy.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review, 67*, 297-308.

Doms, M. E. (1992). *Essays on Capital Equipment and Energy Technology in the Manufacturing Sector*. Ph.D. Dissertation: Univ. of Wisconsin at Madison.

Duranton, G., & Puga, D. (1999). Nursery cities. *London School of Economics, mimeo*.

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

Evenett, S. J., & Keller, W. (2002). On theories explaining the success of the gravity equation. *Journal of Political Economy, 110*.

Feser, E. J., & Bergman, E. M. (2000). National industry cluster templates: a framework for applied regional cluster analysis. *Regional Studies, 34*, 1-20.

Fujita, M. (1988). A monopolistic competition model of spatial agglomeration: differentiated product approach. *Regional Science and Urban Economics, 18*, 87-124.

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

Fujita, M., & Thisse, J.-F. (1996). Economics of agglomeration. *Center for Economic Policy Research Discussion Paper, No 1344.*

Garcia-Mila, T., & McGuire, T. (1993). Industrial mix as a factor in the growth and variability of state's economies. *Regional Science and Urban Economics, 23,* 731-748.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100,* 1126-1152.

Goldar, B. (1997). Econometrics of Indian industry. In K. L. Krishna (Ed.), *Econometric Applications in India* (pp. 53-107). Delhi: Oxford Univ. Press.

Greene, W. (2000). *Econometric Analysis.* Upper Saddle River, NJ: Prentice-Hall.

Hanson, G. H. (1959). How accessibility shapes land use. *Journal of American Institute of Planners, 25,* 73-76.

Henderson, V. (1974). The sizes and types of cities. *American Economic Review, 64,* 640-656.

Henderson, V. (1977). *Economic Thoery and the Cities.* Oxford: Oxford Univ. Press.

Henderson, V. (1988). *Urban Development: Theory, Fact, and Illusion.* New York: Oxford Univ. Press.

Henderson, V. (1999). The effects of urban concentration of economic growth. *NBER Working Paper, 7503.*

Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy, 103,* 1067-1090.

Hotelling, H. (1929). Stability in competition. *Economic Journal, 39,* 41-57.

Infomap, M. (1998). *IndiaMap.* Delhi: ML Infomap Pvt. Ltd.

Isard, W. (1956). *Location and Space-Economy.* Cambridge: MIT Press.

Jacobs, J. (1969). *The Economy of Cities.* New York: Vintage.

Koo, J. (2002). *Agglomeration revisited.* Paper presented at the Southern Regional Science Association Conference, Arlington, VA. In Submission, Annals of Regional Science. .

Krugman, P. (1991a). *Geography and Trade.* Cambridge, Mass.: MIT Press.

Krugman, P. (1991b). Increasing returns and economic geography. *Journal of Political Economy, 99,* 483-499.

Krugman, P. (1993). The Current Case for Industrial Policy. In D. Salvatore (Ed.), *Protectionism and World Welfare*. Cambridge: Cambridge Univ. Press.

Lall, S.V., Shalizi, Z., & Deichmann, U. (forthcoming). Agglomeration economies and productivity in Indian industry. *Journal of Development Economics*. Also, *World Bank Policy Research Working Paper 2663*.

Losch, A. (1956). *The Economics of Location*. New Haven, CT: Yale Univ. Press.

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

Marshall, A. (1919). *Industry and Trade*. London: Macmillan.

Mirachy, W. (1995). *Cities and the Product Cycle*.: M.I.T. Ph. D. Dissertation.

Porter, M. (1990). *The Competitive Advantage of Nations*. London: Macmillan.

Rivera-Batiz, F. (1988). Increasing returns, monopolistic competition, and agglomeration economies in consumption and production. *Regional Science and Urban Economics, 18*, 125-154.

Roth, P. (1994). Missing data: a conceptual review for applied psychologiests. *Personnel Psychology, 47*, 537-560.

Rubin, D. B. (1978). Multiple imputations in sample surveys: a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 20-34.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley & Sons.

Venables, A. (1996). Equilibrium locations of vertically linked industries. *International Economic Review, 49*, 341-359.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics, 80*, 180-207.

von Thunen, J. H. (1966). *Isolated State*. Oxford: Pergamon Press.

Weber, A. (1929). *Theory of the Location of Industries*. Chicago: Univ. of Chicago Press.

Weibull, J. (1976). An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics, 6*, 357-379.

Table 1. Characteristics of firms in the study sectors

Location	Industry	Firms	Employment	Wages/ Employee	Output/ Employee	Value Added/ Employee
Nationwide	All industries	23,201	4,605	60	277	127
	Food processing	4,168	671	47	253	147
	Textiles	3,409	1,111	44	140	76
	Leather	468	79	41	211	135
	Paper products & printing	1,043	129	70	314	204
	Chemicals	2,811	474	83	376	79
	Metals	2,331	410	77	261	114
	Mechanical machinery	1,300	237	78	189	95
	Electrical and electronics	1,267	251	101	344	65
	Other industries	6,404	1,243	54	385	195
Non Urban		8,343	1,494	50	301	126
Non Metro Urban		9,446	1,972	58	235	125
Metro Urban		5,412	1,139	74	320	133

Data for employment, wages/ employee, output/ employee and value added / employee are in thousands. Data Source: ASI 1998-99

Table 2. Spatial concentration

Industry	r	Industry Employment
Food processing	0.031	1,303
Textiles	0.025	1,917
Leather	0.186	159
Paper products & printing	0.013	296
Chemicals	0.011	1,526
Metals	0.088	1,658
Mechanical machinery	0.024	392
Electrical and electronics	0.030	483

Data Source: ASI sample frame, 1998-99
r shows concentration across states
Employment is reported in thousands

36

Table 3. Number of establishments

Industry	Small (0-49)	Medium (50-99)	Large (100+)	Total
Food and beverages	1,821	685	1,498	4,004
Textiles	1,292	406	1,621	3,319
Leather	227	72	144	443
Printing and publishing	663	148	214	1,025
Chemicals	1,549	349	875	2,773
Metals	1,374	291	621	2,286
Mechanical machinery	806	159	318	1,283
Electrical and electronics	711	165	377	1,253
Total	8,443	2,275	5,668	16,386

Data Source: ASI 1998-99

Table 4. Cost elasticities of production factors

	Capital Overall	Small (0-49)	Medium (50-99)	Large (100+)	Labor Overall	Small (0-49)	Medium (50-99)	Large (100+)	Energy Overall	Small (0-49)	Medium (50-99)	Large (100+)	Material Overall	Small (0-49)	Medium (50-99)	Large (100+)
Food and beverages	0.019	0.000	0.000	0.062	0.048	0.004	0.024	0.096	0.128	0.265	0.143	0.028	0.844	0.735	0.860	0.840
Textiles	0.051	0.013	0.021	0.087	0.096	0.029	0.065	0.184	0.141	0.133	0.120	-0.055	0.776	0.890	0.830	0.848
Leather	0.028	0.006			0.079	0.049			0.055	0.182			0.861	0.799		
Printing and publishing	0.082	0.009		0.129	0.139	0.053		0.222	0.318	0.206		0.321	0.692	0.788		0.726
Chemicals	0.069	0.027	0.089	0.103	0.106	0.056	0.124	0.204	0.242	0.263	0.114	0.103	0.716	0.776	0.724	0.683
Metals	0.035	0.004	0.112	0.054	0.092	0.055	0.131	0.134	0.280	0.243	-0.027	0.141	0.618	0.765	0.771	0.804
Mechanical machinery	0.037	0.026		0.084	0.144	0.112		0.192	0.197	0.241		0.096	0.746	0.757		0.689
Electrical and electronics	0.084	0.032		0.169	0.113	0.035		0.182	0.224	0.362		-1.047	0.740	0.650		0.600

Note: Coefficients in bold are significant at 1%, coefficients underlined are significant at 5%.

Table 5. Cost elasticities of economic geography variables

	Access Overall	Small (0-49)	Medium (50-99)	Large (100+)	Emp Overall	Small (0-49)	Medium (50-99)	Large (100+)	IO Link Overall	Small (0-49)	Medium (50-99)	Large (100+)	Diversity Overall	Small (0-49)	Medium (50-99)	Large (100+)
Food and beverages	-0.002	-0.001	-0.001	0.000	0.016	0.000	0.002	0.011	0.006	-0.001	-0.002	0.024	-0.075	0.000	-0.012	-0.067
Textiles	0.004	0.023	0.008	-0.022	0.016	0.004	-0.033	-0.001	-0.002	-0.005	0.037	0.011	-0.102	-0.121	-0.210	-0.005
Leather	0.072	-0.017			0.025	0.005			-0.014	0.010			-0.023	-0.172		
Printing and publishing	0.022	0.005		0.032	0.000	0.012		0.034	-0.009	-0.005		-0.092	-0.062	-0.248		0.0516
Chemicals	-0.016	-0.024	-0.056	-0.044	0.021	0.021	0.059	0.049	0.000	0.003	-0.041	-0.012	-0.076	-0.457	0.042	0.250
Metals	-0.017	-0.008	0.163	-0.012	0.003	0.004	0.137	-0.036	-0.012	0.000	-0.177	0.033	0.003	-0.163	0.603	0.039
Mechanical machinery	-0.047	-0.016		0.091	0.000	0.006		-0.018	-0.001	0.000		-0.026	-0.007	-0.042		0.167
Electrical and electronics	0.008	-0.009		0.035	0.019	0.038		0.035	-0.004	0.004		0.378	-0.162	-0.835		-2.355

Note: Coefficients in bold are significant at 1%, coefficients underlined are significant at 5%.

Table 6. Input Demand Substitution

	Access				Emp				IO Link				Diversity			
	Overall	Small (0-49)	Medium (50-99)	Large (100+)	Overall	Small (0-49)	Medium (50-99)	Large (100+)	Overall	Small (0-49)	Medium (50-99)	Large (100+)	Overall	Small (0-49)	Medium (50-99)	Large (100+)
Food and Beverages																
Capital	-0.050	-0.041	3.670	-0.004	-0.080	0.018	0.119	-2.434	-0.056	-0.019	0.952	0.882	0.249	0.049	-4.751	1.791
Labor	-0.099	-0.056	0.706	0.148	0.000	0.024	0.121	0.173	-0.016	0.017	-0.012	0.113	-0.227	-0.005	-0.767	-0.373
Energy	0.559	0.873	-0.338	0.425	-0.717	0.966	-0.356	0.459	-0.076	0.763	-0.265	0.596	0.508	1.075	-0.334	0.322
Material	-0.037	-0.932	0.342	-0.555	0.057	-0.895	0.376	-0.513	0.012	-0.977	0.266	-0.459	-0.098	-0.852	0.330	-0.594
Textiles																
Capital	-0.484	0.073	-0.012	0.300	-0.078	0.021	0.003	0.020	-0.056	0.121	0.147	-0.062	0.386	-0.008	-0.092	-0.095
Labor	-0.224	-0.228	-0.241	-0.560	-0.096	-0.209	-0.258	-0.454	-0.016	-0.349	-0.164	-0.482	-0.267	-0.403	-0.385	-0.457
Energy	0.716	0.696	1.931	1.550	0.773	0.595	1.792	1.610	-0.076	0.901	1.999	1.596	0.615	0.703	1.678	1.622
Material	-0.548	-0.497	-1.683	-1.132	-0.516	-0.546	-1.794	-1.089	0.012	-0.430	-1.639	-1.092	-0.658	-0.575	-1.936	-1.089
Leather																
Capital	-0.012	-0.157			-0.010	-0.585			0.013	0.608			0.018	-0.050		
Labor	-0.050	0.153			-0.067	-0.020			-0.089	-0.049			-0.137	-0.238		
Energy	-1.161	0.101			-1.125	-0.042			-1.079	-0.004			-1.028	-0.241		
Material	0.957	-0.013			1.170	0.048			1.369	0.059			1.564	-0.120		
Printing and Publishing																
Capital	0.388	-0.570		0.271	0.188	-0.400		0.350	0.052	1.177		-0.130	-0.449	-0.423		-0.224
Labor	-0.373	-0.134		-0.595	-0.337	-0.129		-0.540	-0.404	-0.016		-0.639	-0.422	-0.384		-0.493
Energy	0.338	0.585		0.620	0.180	0.730		0.398	-0.158	0.532		0.978	-0.131	0.469		0.963
Material	0.307	-0.491		0.178	0.326	-0.440		-0.063	0.749	-0.522		0.349	0.476	-0.703		0.271
Chemicals																
Capital	-0.248	-0.085	-0.129	-0.331	-0.147	-0.062	0.077	-0.085	-0.177	-0.085	-0.151	-0.178	-0.187	-0.531	0.029	0.210
Labor	-0.241	-0.536	-0.500	-0.548	-0.038	-0.477	-0.341	-0.281	-0.203	-0.504	-0.480	-0.480	-0.246	-0.963	-0.394	-0.149
Energy	0.487	0.259	0.785	-0.097	-0.332	0.659	1.004	1.039	0.593	0.040	0.795	-0.109	-0.310	0.301	1.090	0.778
Material	0.294	0.225	-0.530	0.313	0.505	0.374	-0.290	0.584	0.193	0.144	-0.475	0.124	0.307	-0.098	-0.270	0.551
Metals																
Capital	0.010	0.108	-0.163	0.139	-0.113	0.455	-0.153	-0.997	0.051	-0.417	-0.483	0.580	0.055	-0.259	0.352	0.462
Labor	0.057	-0.181	0.074	-0.136	0.061	-0.225	-0.001	-0.179	-0.010	-0.106	-0.412	-0.025	-0.042	-0.355	0.426	-0.088
Energy	1.850	0.527	-0.844	0.840	1.940	1.138	-0.961	0.796	1.899	-0.592	-0.917	1.154	2.029	0.118	-0.166	1.087
Material	-2.000	-0.093	1.292	-0.814	-1.934	-0.020	1.202	-0.852	-1.969	-0.210	1.448	-0.657	-1.877	-0.284	2.158	-0.682

Mechanical Machinery

Capital	-0.025	-0.009	-0.032	0.071	0.022	-0.053	**0.163**	**0.132**	-0.048	**0.247**	0.156	0.169
Labor	**-0.204**	**-0.668**	0.645	-0.116	**-0.630**	0.392	**-0.185**	**-0.658**	0.005	**-0.175**	**-0.679**	-0.231
Energy	**1.273**	**1.004**	**-0.916**	**1.429**	**1.206**	**-0.988**	**1.435**	**1.181**	<u>-0.904</u>	**1.587**	**1.347**	-0.723
Material	**-1.419**	**-0.639**	0.748	**-1.303**	**-0.493**	0.766	**-1.300**	**-0.524**	**1.010**	**-1.210**	**-0.411**	**1.259**

Electrical/Electronics

Capital	**-0.235**	**-0.229**	<u>-0.531</u>	-0.131	-0.131	**-0.488**	**-0.215**	<u>-0.199</u>	-0.158	-0.335	**-0.982**	**-2.865**
Labor	**-0.320**	**-0.777**	-0.441	**-0.210**	**-0.714**	**-0.361**	**-0.307**	**-0.761**	-0.095	<u>-0.430</u>	**-1.592**	**-2.803**
Energy	0.225	**1.718**	-0.784	0.023	**1.709**	-0.444	0.093	**1.676**	**-2.819**	-0.260	<u>0.843</u>	**-2.888**
Material	0.681	<u>-0.617</u>	**1.100**	**0.564**	**-0.643**	**1.256**	<u>0.551</u>	**-0.673**	**1.469**	0.396	**-1.523**	-1.158

Note: Coefficients in bold are significant at 1%, coefficients underlined are significant at 5%.

	Title	Author	Date	Contact for paper
WPS3056	Avoiding the Pitfalls in Taxing Financial Intermediation	Patrick Honohan	May 2003	A. Yaptenco 31823
WPS3057	Teaching Adults to Read Better and Faster: Results from an Experiment in Burkina Faso	Helen Abadzi	May 2003	H. Abadzi 80375
WPS3058	Working for God? Evaluating Service Delivery of Religious Not-for-Profit Health Care Providers in Uganda	Ritva Reinikka Jakob Svensson	May 2003	H. Sladovich 37698
WPS3059	Do Capital Flows Respond to Risk and Return?	César Calderón Norman Loayza Luis Servén	May 2003	T. Tourougui 87431
WPS3060	World Market Integration through the Lens of Foreign Direct Investors	Rui Albuquerque Norman Loayza Luis Servén	May 2003	T. Tourougui 87431
WPS3061	Reciprocity in Free Trade Agreements	Caroline Freund	May 2003	P. Flewitt 32724
WPS3062	Labor Effects of Adult Mortality in Tanzanian Households	Kathleen Beegle	May 2003	E. de Castro 89121
WPS3063	Trade Liberalization, Firm Performance, and Labor Market Outcomes in the Developing World: What Can We Learn from Micro-Level Data?	Paolo Epifani	May 2003	P. Flewitt 32724
WPS3064	Trade Policy, Trade Volumes, and Plant-Level Productivity in Colombian Manufacturing Industries	Ana M. Fernandes	May 2003	A. Fernandes 33983
WPS3065	Ghost Doctors: Absenteeism in Bangladeshi Health Facilities	Nazmul Chaudhury Jeffrey S. Hammer	May 2003	H. Sladovich 37698
WPS3066	The Potential Demand for and Strategic Use of an HIV-1 Vaccine in Southern India	Shreelata Rao Seshadri P. Subramaniyam Prabhat Jha	May 2003	H. Sladovich 37698
WPS3067	The Mini-Integrated Macroeconomic Model for Poverty Analysis: A Framework for Analyzing the Unemployment and Poverty Effects of Fiscal and Labor Market Reforms	Pierre-Richard Agénor	May 2003	M. Gosiengfiao 33363
WPS3068	Poverty and Economic Growth in Egypt, 1995–2000	Heba El-laithy Michael Lokshin Arup Banerji	June 2003	P. Sader 33902
WPS3069	International Migration, Remittances, and the Brain Drain: A Study of 24 Labor-Exporting Countries	Richard H. Adams, Jr.	June 2003	N. Obias 31986
WPS3070	Are You Satisfied? Citizen Feedback and Delivery of Urban Services	Uwe Deichmann Somik V. Lall	June 2003	Y. D'Souza 31449