

Economic Geography: Real or Hype?*

Jun Koo
Assistant Professor
Maxine Goodman Levin College of Urban Affairs
Cleveland State University

Somik V. Lall
Senior Economist
Development Research Group
World Bank

Abstract

Economic geography has become a mantra for many economists, geographers, and regional scientists. Many previous studies have tested the importance of economic geography for production activities and found a significant association between them. Most of these studies, however, have not taken into account that economic geography influences location decisions at the firm level. This paper illustrates a potential bias that can arise when firm location choices are not considered in estimating the contribution of economic geography to industry performance. Analysis using microdata of Indian manufacturing firms shows there is an upward bias in the contribution of economic geography to productivity when firm location choices are not considered in the analysis.

World Bank Policy Research Working Paper 3465, December 2004

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

* The research has been partly funded by a World Bank research program grant on “Urbanization and Quality of Life.” The authors can be contacted as follows: Koo – 2121 Euclid Avenue, UR 349, Cleveland State University, Cleveland, OH 44115; Tel 216.687.5597, Fax 216.687.9277, Email jkoo@urban.csuohio.edu; Lall – MC2-621, 1818 H Street NW, World Bank, Washington DC 20433, Email: slall1@worldbank.org

1. Background

The geographic aspect of economic activities has long been of interest to many economists, geographers, planners, and regional scientists. For instance, early location theorists probed the location of industries, land use patterns, and their economic implications (Christaller, 1933; Losch, 1956; von Thünen, 1826; Weber, 1929). Economic geographers have examined how interactions between increasing returns to scale and geographic location lead to a particular distribution pattern of production activities (Krugman, 1980; Pred, 1966). Analytic difficulties in modeling increasing returns to scale, however, marginalized geography in mainstream economic analysis (Krugman, 1991a). As a result, until recently, geography was forgotten in economic research.

Economic geography has since been revived and expanded over the past decade due to advances in mathematical theories that model increasing returns to scale and economies of spatial agglomeration (Dixit & Stiglitz, 1977; Krugman, 1991b). Agglomeration theory, based on such technical development, attributes the geographic concentration of firms to cost-saving externalities. Many recent studies have shown that location is indeed an important factor affecting the economic performance of firms and regions (Beeson, 1987; Feser, 2001; Fogarty & Garofalo, 1988; Henderson, 1986; Moomaw, 1981, 1988). These studies have demonstrated that firms can improve their productivity by locating in large urban areas where similar production activities are concentrated and input factors (e.g., workers) are abundant.

In most empirical models, agglomeration is often treated as a location-specific externality that can occur within the same industry (localization economies) or across all industries as a consequence of the scale of a city or region (urbanization economies) (Feser, 2001; Henderson,

1986; Moomaw, 1988; Nakamura, 1985). Therefore, it varies across industries or locations but is invariant across firms within the same industry or location. Such a specification is meaningful and innovative in that it incorporates spatial aspects of economic activities that have been largely ignored into an economic model. However, it may also introduce a bias arising from a firm's endogenous location decision process. The benefit of locating in a large urban area can be materialized only if a firm makes a location decision accordingly. Firms located in small towns do not benefit from agglomeration economies as much as their counterparts in large cities. Therefore, the agglomeration economies that firms benefit from are a function of firm location choices.

Firms decide their locations to minimize production costs and maximize profit. If a firm is heavily dependent upon natural resources, it will likely locate near those resources to reduce transport costs. On the other hand, if a firm relies heavily on a specialized labor force (i.e., workers with specialized skills), it will likely locate in places where well-educated workers are abundant. Although final location choices of profit-maximizing firms may not be absolute-optimal because firms often have only limited information on markets for factor inputs and other determinants of production costs, they can be at least sub-optimal with respect to cost under constrained information conditions. Accordingly, one can expect firm location choices to follow some systematic patterns. In particular, given that there are centrifugal forces (e.g., competition, congestion, pollution, etc.) as well as well-known centripetal forces in economic geography (i.e., agglomeration economies), more productive firms that can afford a higher cost of doing business are more likely to locate in large urban areas. Firms that rely on out-dated technologies or low-skilled workers may not benefit enough to offset the higher cost of doing business in major cities. In other words, a systematic difference in productivity between firms locating in urban and

rural areas may arise not only from spatial externalities in large cities but also from firms' voluntary choices of production locations.

The discussion thus far raises an interesting issue about the specification of economic geography. It is a proven fact that urban firms are more productive than non-urban firms. Agglomeration theory attributes such a productivity gap to spatial externalities created by well-developed buyer-supplier chains, deep labor pools, and knowledge spillovers in large urban areas (Fujita & Ogawa, 1982; Helsley & Strange, 1990; Venables, 1996). However, the productivity gap may result firm location choices as well. If more productive firms tend to choose urban areas, production function parameter estimates may suffer from a serious selection bias unless the firm location decision process is incorporated into empirical models.

This paper questions the fundamental assumptions of economic geography. If higher productivity of urban firms is indeed associated with individual firms' location decisions, which are developed to minimize their production costs, the implications of economic geography derived from most previous studies can be misleading. When proper consideration is not given to this issue, the effects of economic geography on productivity in many empirical studies are likely to be seriously overestimated. This paper presents a new approach to thinking about the contribution of economic geography to productivity and illustrates this point by estimating simple Cobb-Douglas production functions for 18 2-digit Indian industries as defined by the National Industry Classification (NIC), with and without consideration of firm location choices.

The next section lays out an analytic framework that describes the selectivity issue in the production function estimation and presents an alternative approach that takes into account firm location choices. Section 3 describes the empirical model and hypothesis and Section 4 de-

scribes the data and variables. Section 5 discusses concentration patterns of NIC 2-digit Indian industries and their distributions. Section 6 presents the results, and the last section discusses the implications for research and policy.

2. Modeling a Production Function under Self-Selection

To model a production function under the self-selection process of a location decision, consider a simple production function equation (1) and a location decision equation (2) with a latent variable:

$$O_{ij} = X_{ij}B + u_{ij} \quad (i = 1, 2, 3, \dots, n) \quad (1)$$

$$I_{ij}^* = Z_{ij}R + e_{ij} \quad (j = 1, 2, 3, \dots, m) \quad (2)$$

where O_{ij} is the output of firm i in region j , X_{ij} is a vector of input factors (in log term) utilized by firm i in region j , I_{ij}^* is a latent variable representing firm i 's decision to locate in region j , and Z is a vector of firm and location characteristics that determine the firm location decision process. Since a firm's location decision is an endogenous process influencing agglomeration economies and the firm's productivity, the level of output is conditional upon not only input factors but also location decisions. Therefore, O_{ij} is observed only if firm i chooses to locate in region j , and, consequently, the observed distribution of O_{ij} is truncated. A classic selectivity issue arises as follows:

$$E(O_{ij} | I_{ij} = j) = X_{ij}B + E(u_{ij} | I_{ij} = j) \quad (3)$$

where $I_{ij}=j$ represents a firm's decision to locate in region j . Since $E(u_{ij} | I_{ij} = j) \neq 0$, the OLS estimation of equation (1) will be biased.

Alternatively, following Maddala (1986), a polychotomous-choice model with m categories can be incorporated into the production function framework to correct the self-selection bias. Consider a profit maximizing firm's location decision (subscript i is dropped for simplicity):

$$I = j \text{ iff } I_j^* > \text{Max } I_s^* \quad (4)$$

where $s = 1, 2, 3, \dots, m, j \neq s$. Let

$$\eta_j = \text{Max } I_s^* - e_j \quad (s = 1, 2, 3, \dots, m, j \neq s) \quad (5)$$

Then it follows that

$$I = j \text{ iff } \eta_j < Z_j R \quad (6)$$

Following Domencich and McFadden (1975), the probability for firm i to choose region j is defined as equation (7):

$$\Pr(\eta_j < Z_j R) = \Pr(I = j) = \frac{\exp(Z_j R)}{\sum_s \exp(Z_s R)} \quad (7)$$

Thus, the distribution of η_j can be written as

$$F_j(\eta) = \frac{\exp(\eta)}{\exp(\eta) + \sum_{s=1,2,3,\dots,m(s \neq j)} \exp(Z_s R)} \quad (8)$$

Therefore, for each location choice j , we now have the model $O_{ij} = X_{ij} B + u_{ij}$, where O_{ij} can be observed only if $\eta_j < Z_j R$.

Finally, based on a modified version of Heckman's (1979) two-stage method, we can estimate a production function based on firm location choice behavior. The first stage estimators from equation (2) are obtained by running a modified version of the McFadden's (1974) conditional logit model on firm location choices. After estimating the first stage location choices

specified in equation (7), we can estimate equation (1) with a correction factor derived from the first stage:

$$O_j = X_j B - \sigma_j \rho_j \frac{\phi[\psi_j(Z_j R)]}{F_j(Z_j R)} + v_j \quad (9)$$

where σ_j is the standard deviation of u_{ij} , ρ_j is the correlation coefficient between u_{ij} and e_{ij} , and $\psi_j(Z_j R)$ is the inverse of the standard normal distribution function that transforms non-normal distributions to normal (Lee, 1982).

3. Empirical Model and Hypothesis

To implement the two-stage estimation model proposed in the previous section, we calculate the correction factor as follows. First, a total of 496 districts are categorized as rural, non-metro-urban, and metro-urban areas, and firms are hypothesized to choose their locations among them.¹ We then estimate a conditional logit model by regressing location choices on firm attributes, such as factor intensities, labor productivity, and age, as well as location attributes, such as market access, literacy, and infant mortality rate. The results show that 1) no location-specific attribute significantly affects the odds of choosing a particular location; 2) higher capital intensity increases the odds of locating in metro-urban areas but decreases the odds of locating in non-metro-urban areas; 3) higher labor intensity decreases the odds of locating in non-metro urban or metro-urban areas; 4) higher labor productivity increases the odds of locating in metro-urban areas; and 5) higher age increases the odds of choosing non-metro-urban and metro-urban areas.²

¹ Location categories are defined based on population sizes and our judgment.

² The estimation results are included in Appendix A.

Based on the correction factor calculated from the first-state estimation, a simple Cobb-Douglas production function with economic geography variables are estimated as follows:³

$$\ln O_{ij} = \ln K_{ij} + \ln L_{ij} + \ln E_{ij} + \ln M_{ij} + \sum_e \ln EG_{ej} + C_{ij} \quad (10)$$

where O , K , L , E , and M are output, capital, labor, energy, and material, respectively; C is the location correction factor (i.e., mills ratio) derived from the first-stage location choice model; and EG represents economic geography variables.

We develop economic geography variables based on the new economic geography literature (Fujita, Krugman, & Venables, 1999). First, the transportation infrastructure significantly improves access to markets and inter-regional connectivity. Accordingly, the availability of reliable transportation networks can reduce the unit cost of production and generate consumer surplus, thereby improving productivity and attracting private investment. Two transportation infrastructure-related measures are proposed to capture scale economies from improved market access and transportation networks. Market accessibility reflects the effects of improved access to consumer markets; distance to transport hubs captures the effects of location in transportation networks.

In addition, the model includes industry concentration and urban density variables to capture classic localization and urbanization economies, respectively (Hoover, 1937). Firms located in close proximity to other firms in the same industry often share skilled labor and industry-specific knowledge (i.e., localization economies). They can also benefit from more efficient

³ We also estimated more complicated specifications (e.g., translog). The difference between models with and without the consideration of location choices was not as clear in more complicated models as that simple models. Although more complicated models are still conceptually sound, a large number of parameters may dilute the effects of the location correction factor. Since the purpose of this paper is to illustrate that the importance of economic geography may be exaggerated when firm location choices are not considered, we report the results from simple Cobb-Douglas production function models.

subcontracting and possibilities for collectively lobbying regulators. On the other hand, firms located in large urban areas can benefit from different kinds of sources, such as access to specialized professional services, a large labor pool, and availability of the general infrastructure (i.e., urbanization economies).

If the selectivity issue is indeed relevant, the correction factor is expected to be statistically significant. However, whether incorporating firm location choices into the estimation process will completely wipe out the effects of economic geography is unclear. Although spatial external economies can be offset by the resolved selectivity issue as well as increased costs for labor, land, and transportation, theoretically, economic geography may still play a role (i.e., a smaller role than was believed) in improving firm productivity. Given that more productive firms are likely to locate in large urban areas, we hypothesize that the effects of economic geography variables in the production function estimation are overestimated when firm location choices are not taken into account.

4. Data and Measures

Data. To implement the proposed two-stage estimation model, we use establishment level data from the 1994 Indian Annual Survey of Industries, conducted by the Central Statistical Office of India. The data include various plant level attributes such as output, sales, labor, capital, materials and energy use. These plant level data are supplemented by district and metropolitan level demographic and economic geography variables that are designed to capture scale economies arising from the concentration of economic activities such as improved market access

and localization/urbanization economies. After deleting records that violate simple accounting principles, the total of 47,324 plants are used for the analysis.

Measures. This study measures traditional input and output variables as follows. Output is defined as the ex-factory value of products manufactured for sale during the accounting year. Capital is often measured by perpetual inventory techniques that require continuous observations of the same plant over time. These techniques, however, are difficult to use with micro-level survey data because sample sizes differ by year and a system for tracking firms over time does not exist. Instead, capital is defined as the gross value of the plant and machinery. It includes not only the book value of the installed plant and machinery, but also the approximate value of the rented-in plant and machinery. Doms (1992) demonstrated that it is reasonable to define capital as a gross stock. Labor is defined as the total number of employee mandays worked and paid for by the factory during the accounting year. Energy is measured by the total purchase value of fuels, lubricants, electricity, and water consumed in the production process during the accounting year. Material is measured by the total delivered value of all raw materials, components, chemicals, and packing materials that entered into the production process during the accounting year.

Defining economic geography variables, particularly those related to transportation infrastructure, is not as straightforward as defining traditional input and output variables. In this analysis, we use the transport and market access variables developed in Lall et. al (2004), where access to markets is determined by the distance from and the size of market centers around the plant. Market accessibility is defined as

$$I_i = \sum_j \frac{S_j}{d_{ij}^b} \quad (11)$$

where I_i is the accessibility indicator estimated for location i , S_j is a size indicator at destination j (e.g., population, purchasing power, or employment), d_{ij} is a measure of distance between origin i and destination j , and b describes how increasing distance reduces the expected level of interaction. The measure is constructed based on the Indian road network and urban population centers. Lall et. al (2004) also calculated distances (measured by travel times) between district centroids and transport hubs to examine if a short travel time to transport hubs has external economies above and beyond the effects of market accessibility.

At the industry level, a simple location quotient (LQ) is used to measure localization economies. In addition, this study uses urban population density (i.e., the ratio of the urban population to the urban area of the district) as an indicator for urban scale economies. While many other studies have used urban sizes as a proxy for urbanization economies, we use density because it better reflects spatial concentration.

5. Spatial Industrial Concentration in India

The essence of economic geography is the spatial concentration of economic activities and subsequent economic benefits. Therefore, examining spatial concentration patterns of firms is the first necessary step when investigating economic geography. This section presents a brief overview of spatial industrial concentration in India. We examined spatial concentration patterns of 18 NIC 2-digit Indian industries using a concentration measure that Ellison and Glaeser (1997) recently proposed:

$$r = \frac{\sum_{i=1}^M (s_i - x_i)^2 - (1 - \sum_i^M x_i)H}{(1 - \sum_i^M x_i)(1 - H)} \quad (12)$$

where s_i is region i 's share of the study industry, x_i is the regional share of total employment, and H is the Herfindahl industry plant size distribution index, $H = \sum_{j=1}^N z_j^2$.

The Ellison-Glaeser (EG) index has several advantages over other widely used concentration indexes, such as location quotients (LQ) and Gini coefficients. First, the index is developed based on an explicit micro theory because it is derived from firm location choices. Second, the index takes on a value of zero when plant location distribution patterns are random (as opposed to uniform). Therefore, it captures agglomeration above and beyond what we would observe if firm location decisions were random. Third, the index is designed to make comparisons across industries, countries, and over time.

[Table 1 Here]

We calculate the raw concentration measure G , Herfindahl index H , and EG index r for 18 NIC 2-digit Indian industries. Following Ellison and Glaeser's definition of concentration ($r < 0.02$: not very localized, $0.02 \leq r \leq 0.05$: intermediate, and $r > 0.05$: highly localized), jute textile, beverages, leather/leather products, miscellaneous food products n.e.c., wood/wood products, textile products, and wool/silk products show very high levels of local concentration, whereas non-metallic mineral, transport equipment/parts, machinery other than transport/electronic/electrical, electronic/electrical machinery/parts/apparatus, rubber/petroleum/coal products, metal, and paper/paper products are hardly localized. The results indicate that more resource-intensive industries tend to be more locally concentrated. Overall, spatial industrial distribution patterns in India resemble the concentration patterns of the U.S. manufacturing industries that Ellison and Glaeser investigated.

We then examine labor productivity in rural, nonmetro-urban, and metro-urban areas. A simple comparison of productivity does not prove any causal relationship between economic geography and productivity differences. It is, however, meaningful since it can highlight important characteristics of firms located in different areas, which might result from location choices. Table 2 illustrates that there is a noticeable difference in labor productivity among firms in rural, nonmetro-urban, and metro-urban areas. Firms in large urban areas are substantially more productive than those in rural areas. The difference might be an outcome of economic geography, firm location choices, or both.

[Table 2 Here]

6. Results

To illustrate a potential bias created by the firm location decision process, we estimate two sets of Cobb-Douglas production functions for 18 NIC 2-digit Indian industries: one with the location correction factor derived based on firm location choices and the other without it. For both cases, we run simple OLS models with and without regularity restrictions (i.e., monotonicity and quasiconcavity). Regularity restrictions do not make any substantial difference in overall results. Therefore, this section discusses results from models with regularity restrictions.

A major difference between this paper and others is the inclusion of the location correction factor in the production function estimation, which will demonstrate a potential selection bias arising from firm location choices. The significance level of the correction factor suggests whether the two-stage estimation process that takes into account firm location choices is indeed necessary. If the correction factor is not statistically significant, firm location choices will not

create any estimation bias. This implies that firms make their location decisions randomly. It is often the case in developing countries where information on the market is limited. In other words, individual firms may make rational decisions with limited information. The collective firm location patterns, however, can be close to random. Therefore, a comparison between the corrected model (with the correction factor) and the uncorrected model (without the correction factor) can illustrate a potential selection bias caused by firm location choices.

The correction factor is statistically significant in 15 out of 18 NIC 2-digit Indian industries, indicating a strong selection bias. Among economic geography variables, location quotient and urban density, which represent localization and urbanization economies, show mixed signs. Table 3 shows that, in both corrected and uncorrected models, the location quotient affects output levels negatively in six industries (miscellaneous food products n.e.c., non-metallic mineral products, metal products, textile products, wood/wood products, paper/paper products) and positively in three industries (wood/silk textiles, transport equipment/parts, and leather/leather products). In addition, urban density affects output levels negatively in five industries (food products, miscellaneous food products n.e.c., chemical/chemical products, wood/silk textiles, and transport equipment/parts) and positively in two industries (jute textile and textile products). This implies that centrifugal forces as well as centripetal forces of economic geography are in place. Firms are expected to benefit from spatial scale externalities arising from buyer-supplier linkages, a deep labor pool, knowledge spillovers, and the availability of specialized services, and a general infrastructure. On the other hand, a significant concentration of economic activities can also cause negative externalities, such as competition, congestion, and pollution that will increase the cost of doing business.

[Table 3 Here]

The two transportation-related economic geography variables show clearer patterns of association with output levels. In uncorrected models, market access significantly increases output levels in 11 industries (miscellaneous food products n.e.c., beverages, chemical/chemical products, rubber/petroleum/coal products, wood/silk textiles, basic metals/alloys, machinery other than transport/electronic/electrical, electronic/electrical machinery/parts/apparatus, textile products, paper/paper products, leather/leather products); distance to transport hubs significantly decreases output levels in 12 industries (food products, chemical/chemical products, rubber/petroleum/coal products, cotton textiles, wool/silk textiles, basic metals/alloys, metal products, machinery other than transport/electronic/electrical, electronic/electrical machinery/parts/apparatus, transport equipment/parts, paper/paper products, and leather/leather products).

An interesting pattern emerges when the correction factor is added to the estimation. Market access loses its statistical significance in five industries (chemical/chemical products, rubber/petroleum/coal products, electronic/electrical machinery/parts/apparatus, paper/paper products, and leather/leather products), and distance to transport hubs loses statistical significance in two industries (chemical/chemical products, leather/leather products). This implies that the traditional production function estimation, which ignores firm location choices, can create a bias and wrongly reject the null hypothesis of parameter estimates. In addition, the results also suggest that the importance of transportation infrastructure in particular may not be as critical as was believed after firm location choices are taken into account.

As far as the magnitude of parameters is concerned, economic geography variables in uncorrected models have a stronger influence on output levels than those in corrected models. In other words, the absence of the correction factor tends to inflate parameter estimates of the economic geography variables. In particular, when the correction factor is not included, the influence of market access and distance to transport hubs is exaggerated in 11 and 12 out of 18 industries, respectively. When these two variables are statistically significant, they are always overestimated without the correction factor. If we only consider industries with statistically significant correction factors, the importance of market access is overestimated in 10 out of 15 industries, and that of distance to transport hubs is also inflated in 12 out of 15 industries.

The results thus far indicate that the importance of economic geography, particularly the benefit of transportation infrastructure to productivity, is somewhat oversold. Estimates for scale externalities from the transportation infrastructure can be more significantly biased by firm location choices than those for localization and urbanization economies. The transportation infrastructure is still, however, an important determinant of productivity for many firms and industries since market access and distance to transport hubs still play strong roles in production activities in six and ten industries, respectively, even after controlling for firm location choices. In sum, economic geography may not be hype, but its effects are not as real as typically believed.

7. Conclusion

Economic geography has become a mantra for many economists, geographers, and regional scientists. Many previous studies have tested the importance of economic geography for production activities and found a significant association between them. Methodologically, however, they have not taken into account that economic geography influences firm location choices.

In other words, most previous research did not acknowledge that spatial scale economies in large urban areas are materialized only after firms make their location decisions accordingly. When a contingent nature of economic geography is ignored, the validity of empirical findings can be seriously questioned.

This paper proposes a new approach to thinking about economic geography and illustrates a potential bias that can arise when firm location choices are not considered as part of economic geography. An analysis using microdata of Indian manufacturing firms shows that when firm location choices are not given proper consideration, the role of economic geography can be overemphasized. This is particularly true for transportation infrastructure. The results indicate that the importance of market access and distance to transport hubs is exaggerated in many industries.

Economic geography still matters to many firms and industries even after firm location choices are taken into account as part of economic geography. Its magnitude, however, is not as significant as has been believed. Therefore, policymakers need to exercise caution when interpreting results from previous research and applying them to future regional development strategies.

Reference

- Beeson, P. (1987). Total factor productivity growth and agglomeration economies in manufacturing, 1959-73. *Journal of Regional Science*, 27, 183-199.
- Christaller, W. (1933). *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer.
- Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67, 297-308.
- Domencich, T., & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland.
- Doms, M. E. (1992). *Essays on Capital Equipment and Energy Technology in the Manufacturing Sector*. Ph.D. Dissertation: Univ. of Wisconsin at Madison.
- Ellison, G., & Glaesar, E. L. (1997). Geographic concentration in U.S. manufacturing: a dashboard approach. *Journal of Political Economy*, 105(5), 889-927.
- Feser, E. (2001). A flexible test for agglomeration economies in two US manufacturing industries. *Regional Science and Urban Economics*, 31, 1-19.
- Fogarty, M., & Garofalo, G. (1988). Urban spatial structure and productivity growth in the manufacturing sector of cities. *Journal of Urban Economics*, 23, 60-70.
- Fujita, M., Krugman, P., & Venables, A. (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.
- Fujita, M., & Ogawa, H. (1982). Multiple equilibrium and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12, 161-196.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153-161.
- Helsley, R., & Strange, W. (1990). Matching and agglomeration in a system of cities. *Regional Science and Urban Economics*, 20, 189-212.
- Henderson, V. (1986). Efficiency of resource usage and city size. *Journal of Urban Economics*, 19, 47-70.
- Hoover, E. M. (1937). *Location Theory and the Shoe and Leather Industries*. Cambridge, MA: Harvard Univ. Press.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70, 950-959.
- Krugman, P. (1991a). *Geography and Trade*. Cambridge, Mass.: MIT Press.
- Krugman, P. (1991b). Increasing returns and economic geography. *Journal of Political Economy*, 99, 183-199.
- Lall, S., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. *Journal of Development Economics*, 73, 643-673.
- Lee, L. F. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49, 355-372.
- Losch, A. (1956). *The Economics of Location*. New Haven, CT: Yale Univ. Press.
- Maddala, G. S. (1986). *Limited Dependent and Qualitative Variables in Econometrics*: Cambridge Univ. Press.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.
- Moomaw, R. L. (1981). Productivity and city size: A critique of the evidence. *Quarterly Journal of Economics*, 96, 675-688.

- Moomaw, R. L. (1988). Agglomeration economies: localization or urbanization? *Urban Studies*, 25, 150-161.
- Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries, a case of Japanese cities. *Journal of Urban Economics*, 17, 108-124.
- Pred, A. (1966). *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800-1914: Interpretive and Theoretical Essays*. Cambridge, MA: MIT Press.
- Venables, A. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, 49, 341-359.
- von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg: F. Perthes.
- Weber, A. (1929). *Theory of the Location of Industries*. Chicago: Univ. of Chicago Press.

[Table 1] Concentration of Indian Industries

Industry	NIC Code	No. of States	G	H	r
<i>Jute Textiles</i>	25	12	0.548	0.021	0.570
<i>Beverages</i>	22	23	0.313	0.019	0.329
<i>Leather and Leather Products</i>	29	17	0.143	0.012	0.146
<i>Miscellaneous Food Products, n.e.c.</i>	21	24	0.092	0.003	0.098
<i>Wood and Wood Products</i>	27	26	0.079	0.007	0.080
<i>Textile Products</i>	26	20	0.066	0.002	0.070
<i>Wool and Silk Textiles</i>	24	20	0.058	0.006	0.058
<i>Food Products</i>	20	26	0.043	0.001	0.046
<i>Basic Metals and Alloys</i>	33	24	0.053	0.020	0.038
<i>Cotton Textiles</i>	23	21	0.029	0.002	0.030
<i>Chemicals and Chemical Products</i>	30	24	0.027	0.002	0.027
Non-Metallic Mineral Products	32	26	0.019	0.001	0.019
Transport Equipment and Parts	37	22	0.025	0.009	0.018
Machinery other than Transport/Electronic/Electrical	35	22	0.018	0.006	0.013
Electronic and Electrical Machinery, Parts, and Apparatus	36	24	0.018	0.009	0.010
Rubber, Petroleum and Coal Products	31	24	0.011	0.005	0.007
Metal Products	34	27	0.007	0.004	0.002
Paper and Paper Products	28	25	0.006	0.004	0.002
Mean			0.083	0.008	0.083

Source: Annual Survey of Indian Industries

[Table 2] Location and Productivity

	No. of Firms	Labor Productivity
Rural	12,378	1,022.7
Non-metro Urban	24,691	1,163.6
Metro Urban	10,255	1,391.2
Total	47,324	1,176.0

Source: Annual Survey of Indian Industries

[Table 3] Cobb-Douglas Production Function Estimation with Economic Geography Variables*

Food Products					Chemical and Chemical Products				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	4.079	0.134	3.991	0.130	Intercept	2.524	0.184	2.063	0.169
Capital	0.090	0.006	0.092	0.006	Capital	0.074	0.006	0.079	0.006
Labor	0.250	0.009	0.248	0.009	Labor	0.256	0.010	0.250	0.010
Energy	0.198	0.008	0.199	0.008	Energy	0.164	0.008	0.165	0.008
Material	0.431	0.003	0.431	0.003	Material	0.529	0.006	0.528	0.006
LQ	-0.001	0.005	-0.002	0.005	LQ	-0.004	0.006	-0.006	0.006
Density	-0.055	0.004	-0.055	0.004	Density	-0.013	0.006	-0.011	0.006
Access	-0.021	0.013	-0.017	0.013	Access	0.019	0.017	0.049	0.017
Hub	-0.018	0.003	-0.020	0.003	Hub	-0.005	0.003	-0.009	0.003
Correction	0.043	0.016			Correction	0.125	0.020		
Miscellaneous Food Products, n.e.c.					Rubber, Petroleum and Coal Products				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	3.440	0.197	2.991	0.195	Intercept	2.107	0.212	1.773	0.189
Capital	-0.009	0.008	-0.005	0.008	Capital	0.080	0.008	0.085	0.008
Labor	0.410	0.009	0.413	0.009	Labor	0.336	0.014	0.334	0.014
Energy	0.161	0.008	0.159	0.009	Energy	0.180	0.011	0.178	0.011
Material	0.442	0.004	0.442	0.004	Material	0.466	0.007	0.465	0.007
LQ	-0.034	0.007	-0.041	0.007	LQ	0.007	0.006	0.006	0.006
Density	-0.033	0.005	-0.038	0.005	Density	-0.003	0.007	-0.002	0.007
Access	0.077	0.017	0.103	0.017	Access	0.033	0.020	0.057	0.019
Hub	0.006	0.005	-0.002	0.005	Hub	-0.011	0.003	-0.014	0.003
Correction	0.224	0.022			Correction	0.075	0.022		
Beverages					Non-metallic Mineral Products				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	2.689	0.434	2.214	0.386	Intercept	2.447	0.132	2.390	0.126
Capital	0.036	0.012	0.037	0.012	Capital	0.039	0.005	0.040	0.005
Labor	0.328	0.019	0.326	0.019	Labor	0.321	0.010	0.321	0.010
Energy	0.170	0.019	0.172	0.019	Energy	0.249	0.006	0.248	0.006
Material	0.446	0.012	0.444	0.012	Material	0.420	0.005	0.420	0.005
LQ	0.009	0.014	0.011	0.014	LQ	-0.013	0.004	-0.013	0.004
Density	0.014	0.015	0.019	0.015	Density	0.008	0.004	0.007	0.004
Access	0.082	0.041	0.117	0.038	Access	-0.001	0.013	0.002	0.013
Hub	0.009	0.010	0.005	0.010	Hub	0.017	0.004	0.016	0.004
Correction	0.122	0.051			Correction	0.025	0.018		

* Bold represents significant economic geography variables at <0.05; grey scale represents potentially overestimated economic geography variables.

Cotton Textiles

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	4.375	0.187	4.244	0.182
Capital	0.052	0.008	0.054	0.008
Labor	0.219	0.012	0.219	0.012
Energy	0.196	0.010	0.195	0.010
Material	0.431	0.004	0.430	0.004
LQ	0.006	0.007	0.007	0.007
Density	-0.032	0.009	-0.028	0.009
Access	0.001	0.018	0.003	0.018
Hub	-0.019	0.006	-0.021	0.006
Correction	0.070	0.024		

Wool and Silk Textiles

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	3.868	0.253	3.474	0.245
Capital	0.098	0.010	0.106	0.010
Labor	0.204	0.015	0.209	0.015
Energy	0.138	0.013	0.132	0.013
Material	0.462	0.006	0.461	0.006
LQ	0.021	0.006	0.018	0.006
Density	-0.031	0.011	-0.023	0.011
Access	0.052	0.027	0.066	0.027
Hub	-0.019	0.006	-0.030	0.005
Correction	0.182	0.033		

Jute Textiles

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	1.267	1.115	1.316	0.987
Capital	0.026	0.028	0.026	0.028
Labor	0.255	0.054	0.254	0.054
Energy	0.191	0.044	0.191	0.044
Material	0.472	0.024	0.473	0.024
LQ	0.046	0.029	0.045	0.027
Density	0.126	0.032	0.126	0.032
Access	0.122	0.104	0.119	0.098
Hub	-0.034	0.019	-0.034	0.018
Correction	-0.010	0.110		

Basic Metals and Alloys

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	2.195	0.156	1.809	0.149
Capital	0.088	0.006	0.095	0.006
Labor	0.192	0.012	0.188	0.012
Energy	0.170	0.008	0.171	0.009
Material	0.529	0.006	0.529	0.006
LQ	0.010	0.005	0.009	0.005
Density	0.002	0.006	0.005	0.006
Access	0.062	0.014	0.080	0.014
Hub	-0.007	0.003	-0.012	0.003
Correction	0.132	0.018		

Metal Products

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	2.707	0.179	2.560	0.160
Capital	0.086	0.007	0.087	0.007
Labor	0.295	0.011	0.296	0.011
Energy	0.142	0.008	0.141	0.008
Material	0.493	0.006	0.492	0.006
LQ	-0.024	0.006	-0.025	0.006
Density	-0.001	0.006	0.001	0.006
Access	0.000	0.017	0.010	0.016
Hub	-0.006	0.003	-0.008	0.003
Correction	0.038	0.021		

Machine other than Transport/Electronic/Electrical

Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr
Intercept	2.328	0.158	2.084	0.144
Capital	0.055	0.006	0.057	0.006
Labor	0.327	0.011	0.329	0.011
Energy	0.145	0.009	0.142	0.009
Material	0.515	0.005	0.513	0.005
LQ	-0.002	0.005	-0.004	0.005
Density	-0.009	0.006	-0.007	0.006
Access	0.037	0.016	0.055	0.016
Hub	-0.017	0.003	-0.020	0.003
Correction	0.066	0.018		

* Bold represents significant economic geography variables at <0.05; grey scale represents potentially overestimated economic geography variables.

Textile Products					Electronic and Electrical Machinery, Parts, and Apparatus				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	0.977	0.350	0.835	0.325	Intercept	2.453	0.227	2.046	0.187
Capital	0.028	0.011	0.028	0.011	Capital	0.047	0.008	0.050	0.007
Labor	0.362	0.016	0.366	0.015	Labor	0.317	0.013	0.320	0.013
Energy	0.196	0.015	0.196	0.015	Energy	0.148	0.011	0.145	0.011
Material	0.421	0.007	0.419	0.007	Material	0.534	0.007	0.532	0.007
LQ	-0.024	0.011	-0.026	0.011	LQ	-0.002	0.006	-0.003	0.006
Density	0.027	0.016	0.033	0.015	Density	-0.002	0.008	0.000	0.007
Access	0.245	0.038	0.250	0.038	Access	0.011	0.024	0.043	0.021
Hub	-0.002	0.006	-0.003	0.006	Hub	-0.010	0.004	-0.013	0.004
Correction	0.043	0.040			Correction	0.074	0.023		
Wood and Wood Products					Transport Equipment and Parts				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	4.086	0.279	3.632	0.270	Intercept	3.459	0.269	2.902	0.244
Capital	0.000	0.010	0.001	0.010	Capital	0.017	0.011	0.026	0.010
Labor	0.330	0.018	0.332	0.018	Labor	0.368	0.015	0.377	0.015
Energy	0.287	0.014	0.283	0.014	Energy	0.120	0.014	0.111	0.014
Material	0.361	0.006	0.360	0.006	Material	0.509	0.008	0.506	0.008
LQ	-0.019	0.009	-0.015	0.009	LQ	0.022	0.008	0.021	0.008
Density	0.007	0.007	0.011	0.007	Density	-0.038	0.011	-0.038	0.011
Access	-0.040	0.024	-0.011	0.024	Access	0.010	0.029	0.042	0.028
Hub	0.003	0.005	-0.002	0.005	Hub	-0.021	0.005	-0.027	0.005
Correction	0.173	0.031			Correction	0.141	0.030		
Paper and Paper Products					Leather and Leather Products				
Variable	Corrected Model		Uncorrected Model		Variable	Corrected Model		Uncorrected Model	
	Estimate	StdErr	Estimate	StdErr		Estimate	StdErr	Estimate	StdErr
Intercept	2.668	0.227	2.449	0.201	Intercept	3.143	0.758	2.030	0.647
Capital	0.076	0.008	0.077	0.008	Capital	0.002	0.019	0.010	0.018
Labor	0.339	0.013	0.340	0.013	Labor	0.382	0.027	0.387	0.027
Energy	0.130	0.009	0.129	0.009	Energy	0.239	0.024	0.225	0.023
Material	0.470	0.007	0.469	0.007	Material	0.395	0.010	0.394	0.010
LQ	-0.040	0.007	-0.039	0.007	LQ	0.029	0.014	0.029	0.014
Density	0.007	0.007	0.010	0.007	Density	0.023	0.019	0.007	0.018
Access	0.023	0.021	0.039	0.020	Access	0.037	0.081	0.153	0.070
Hub	-0.007	0.003	-0.009	0.003	Hub	-0.007	0.010	-0.020	0.009
Correction	0.053	0.025			Correction	0.179	0.064		

* Bold represents significant economic geography variables at <0.05; grey scale represents potentially overestimated economic geography variables.

Appendix A. Location Selection Model Estimation

Variable	Coefficient	Hazard Ratio
Non-metro-urban	0.871150*	2.390
Metro-urban	-0.241080*	0.790
Market Access	0.000001	1.000
Literacy	-0.000280	1.000
Infant Mortality	0.006510	1.007
Capital intensity*Non-metro urban	-0.466980*	0.627
Capital intensity*Metro-urban	0.722100*	2.059
Labor intensity*Non-metro-urban	-0.756240*	0.469
Labor intensity*Metro-urban	-0.336060*	0.715
Labor productivity*Non-metro-urban	0.000010	1.001
Labor productivity*Metro-urban	0.000063*	1.001
Age*Non-metro-urban	0.000942*	1.001
Age*Metro-urban	0.000869*	1.001

* Significant at <0.05