Staff Paper P94-14 June 1994

STAFF PAPER SERIES

The Role of Agglomeration Economies in Firm Location: A Review of the Literature

by

Anne Selting

Christopher Allanach

Scott Loveridge

DEPARTMENT OF AGRICULTURAL AND APPLIED ECONOMICS

COLLEGE OF AGRICULTURE

UNIVERSITY OF MINNESOTA

Staff Paper P94-14

June 1994

The Role of Agglomeration Economies in Firm Location: A Review of the Literature

by

Anne Selting

Christopher Allanach

Scott Loveridge

The authors are Ph.D. candidate, Dept. Agricultural Economics, University of California, Davis; and Graduate Research Assistant and Assistant Professor, Dept. Agricultural and Applied Economics, University of Minnesota.

The authors thank Wilbur Maki and George Morse for comments on an earlier version of this paper.

Information on other titles in this series may be obtained from: Waite Library, University of Minnesota, Department of Agricultural and Applied Economics, 232 Classroom Office Building, 1994 Buford Ave., St. Paul, MN 55108, USA.

I. The Firm's Location Decision Reviewed

If firms were animals rather than economic entities, a behavioral scientist trying to describe their traits would observe that firms tend to be found in herds and usually migrate toward the largest watering holes. Acknowledging that form follows function, the compelling question is why firms so consistently cluster in and around metropolitan areas. It is difficult to decipher and quantify the intrinsic benefits that reward this adaptation to the economic environment.

Economic analysis offers some insight into how cities form. Nascent location theory emphasized the physical attributes of an area and concluded that settlements arise in places that have comparative advantages inherent in a site. Examples include fertile land, port access, or abundant natural resources. While studies such as the pioneering work of Adna Weber in 1899 provide historical insight into the development of urban centers, the original attractiveness of a city offers little insight into why the area continues to grow or why firms continue to move into its perimeters. (Hoover and Giarratani, 1984).

A more precise explanation of firm location developed around the importance of transportation costs. Optimal sites were assumed to minimize the combined cost of shipping raw materials to the production site and transporting the finished product to the market (Weber, 1956). Later theory focused on the importance of substitutability of input costs, recognizing there may not be a single optimal location and that higher transport expenses will be accepted if an area has compensating low-cost inputs, such as a cheap labor supply (Moses, 1958). Recent work has articulated that overall cost minimization may not be the sole objective in location decisions (Blair and Premus, 1987). Modern extensions have considered the firm's site selection as a complex and multi-faceted process.

Recognizing the complexity of firm location decisions is important but does not lead to the discovery of the underlying factors driving firm location behavior. Without a better understanding of the fundamental calculus of these decisions, the selection of one site over another appears to be random, chaotic and individual to the circumstances of each firm. A more rigorous explanation has been sought to answer three central questions of firm location: why firms tend to locate together, why firms tend to cluster in urban areas and why there is currently a trend for firms to locate in suburban (ring) areas.

Population and industry statistics illustrate these tendencies. Two major trends in the growth of U.S. cities are clear. The first is the widespread and continued migration to city areas by both population and business. U.S. out-migration from rural areas has been steady for several decades. From 1950 to 1990, the percentage of population located in or around urban areas has grown steadily from 56.3 percent to over 77 percent, and the bulk of population and industry is centered around SMSAs[1] (Statistical Abstract of the United States, 1990).[2] Heavy industry concentration in urban areas ensures that the majority of business is centered around cities. In Minnesota, for example, nearly 60 percent of all manufacturing firms in the state are located within a seven-county metropolitan area, despite the relatively higher costs of land and labor (Minnesota Average Covered Employment and Wages, Department of Jobs and

[1]SMSA stands for standard metropolitan statistical area and is the unit of measurement for cities as defined by the U.S. Office of Management and Budget.

[2]This phenomenon is not limited to the U.S., nor to developed countries. Globally, half of the world's population will live in cities by the year 2000 (UN Chronicle, 1986).

Training, 1991, p. 3 and 29).[3] Similar concentrations of manufacturing can be found in the U.S. as a whole and other developed nations (Krugman, 1991).

While movement toward city areas continues, population and industry location within cities is changing. The second significant trend is the rapid growth in business and population on the edge of cities, the so-called "ring effect." Suburban areas have become active centers of industry, and relatively underdeveloped rural areas are being absorbed by expanding suburbs. As one urban geographer notes, "The growth of suburban downtown represented the most important structural change in the metropolitan area of the late twentieth century" (Hartshorn, 1992, p. 400).

The birth of the suburb as an important center of economic activity has had a substantial impact on business location patterns. Since 1970, the share of Fortune 500 service and industrial firms selecting a suburban location for its headquarters has shifted significantly away from downtown urban locations. Major cities such as Chicago, New York (Manhattan), Los Angeles and Philadelphia, have lost sizable numbers of corporate clients to the suburbs. Three American cities, Chicago, Los Angeles and New York, now have equal numbers of suburban and downtown corporations. (Stanback, 1991). In Minneapolis, 98 percent of all job growth in 1980-90 occurred in the suburbs, versus 4 percent growth in downtown employment.[4] The suburbs seem to be where much of business is finding its greatest locational advantage. Only businesses highly dependent on information-intensive, one-on-one interactions still find the downtown area to contain a significant locational advantage (Maki, 1992).

These trends are curious because they are occurring in an era in which many of the amenities important to business are no longer exclusive to cities. Strongly integrated air, road and rail transport have eliminated the unreliability of doing business over long distances and have effectively shortened the span between supply and demand markets. Advanced communications are making the delivery of information nearly instantaneous. Even face-to-face contact is possible via interactive video. Finally, markets have rapidly become global. Thus, the "end of the line" for products is no longer the nearest large city. This tends to blur the original advantages of locating production in the city to be close to customers.

In contrast, the costs of city life and work are increased crime, pollution, congestion, higher living costs and longer commutes for city residents and higher fixed costs for business. Costs that impact society are also recognized. In the midwest, for example, suburban sprawl has brought on complaints that scenic rural country and its area farms and natural habitat are being swallowed. Fearing unmitigated encroachment by other cities or uncontrolled expansion of its own boundaries, many communities have enacted growth control policies. Against these growing costs, firms still tend to prefer city locations and gravitate toward other firms in the same or related industry. If it is assumed that a firm rationally selects its site, the decision to locate in a city area signals that there are unique economic advantages of this location. The key research concern is to isolate the nature of these advantages. While no single theory can satisfactorily address the nuances involved in all firm location decisions, the most consistent theoretical explanation advanced to explain firm location choices is **agglomeration economies**. The theory of agglomeration economies includes two main benefits that a firm may incur as a result of its

---

[3]While there is a substantial representation of manufacturing in urban areas, the growth rate in U.S. manufacturing activities in non-metropolitan areas has also been significant (Dorf and Emerson, 1978). Minneapolis, for example, lost 25 percent of its manufacturing jobs from 1980 to 1991 (Star Tribune, 6/28/93).

[4]The study looked at aggregate job movement in the Twin Cities. St. Paul experienced a 2 percent loss in jobs.

location: (1) the advantages associated with locating near other firms within the same industry, and (2) the advantages that result from locating near an area of dense population. The first advantage is related to the interdependence of firms within an industry. For example, the need for specialized inputs could make locating near certain firms cost effective. The second type of benefit includes savings that relate to market accessibility. Cost reductions in transportation and communication can enhance the competitive position of a firm.

II. Agglomeration Economies: A Working Definition

Arriving at a satisfying definition for agglomeration economies is a muddy task. Pascal and McCall comment, "Although the dictionary treats agglomeration as a virtual synonym of clustering, the term seems to carry a heavier, if amorphous, burden in economists' writings. . . ." (1980, p. 383). Several definitions are offered here to build an understanding of the meaning of agglomeration in an economist's vernacular. In its simplest sense, agglomeration economies "exist when resources are more productive in large cities than in small ones" (Helsley and Strange, 1991, p. 96). Kaldor (1970) defines agglomeration economies as:

> Nothing else but the existence of increasing returns to scale - using that term in the broadest sense - in processing activities. These are not just the economies of large-scale production, commonly considered, but the cumulative advantages accruing from the growth of industry itself (p. 340).

An alternative definition is given by Weber (1956):

> An agglomerative factor, for purposes of our discussion is an "advantage" or cheapening of production or marketing which results from the fact that production is carried on at some considerable extent at ***one*** place, while a deglomerative factor is cheapening of production which results from the decentralization of production (p. 126).

Thus, agglomeration confers advantages to firms by either stimulating consumption (such as providing better access to markets) **or** decreasing the costs of production.

Agglomeration economies are a form of external economies, and this distinction is important. The benefits of agglomeration are not under the control of the firm. As Meyer describes it, if firm A experiences a change in its cost or production function because firms B and C locate in close proximity, the benefits accruing to A result without B and C provoking the circumstance. Firms B and C are not compensated for contributing to A's enterprise, nor can A influence B or C's actions. As Meyer comments, "the firm cannot itself create external economies" (1977, p. 81) but these inadvertent benefits can substantially impact the operations of the affected firms.

The exact source of agglomeration economies is a point of contention in the literature. Nourse (1968) identifies four components of agglomeration economies generally accepted as important. They are: transfer economies, internal economies of scale, economies of scale external to the firm and economies of scale external to the industry. The positive influence that each has on a firm's cost or production functions is addressed below.

A. Transfer Economies

This refers to the savings in transport costs firms receive by locating close to other businesses. The importance of these savings is positively related to the bulk of the firm's inputs and outputs. Based on the nature of its product, the firm decides whether it is more important to locate close to the materials used in production or to the market for its output. If firms cluster, they can supply inputs to one another and thus save on the cost of transporting materials from distant locales. This lowers the average total cost curve of each firm. Savings on information are also a form of transfer economies and are positively related to the complexity of information. The need for face-to-face contact may increase the advantages of locating close to other firms. Complex information can be exchanged at a lower cost and with less travel time. Transfer economies may also occur on the output side, saving the consumer time.

B. Internal Economies of Scale

Economies of scale are the movement down the long-run average cost of a firm caused by expanding the level of its output and accrue as the firm increases its scale of production. These cost savings can be brought about by a variety of factors including division of labor, integration of processes, economies of increased dimensions, economies of specialization, technology, learning, superior technique and organization of production (Meyer, 1977). Clustering can positively influence a firm's internal economies of scale. For example, if firm B locates in firm A's vicinity and requires firm A's output for its own production process, firm A is able to expand its market. This enlargement of production in firm A can assist the plant in attaining a lower cost per unit because of the scale economies involved in production. Meyer (1977) warns that to preserve the external nature of agglomeration, it is important to avoid labelling all internal economies of scale as some sort of agglomeration benefits. He provides an example to make the distinction. A firm owner must decide whether to locate three production plants together or separately. If the plants are all placed in the same area and the firm consequently experiences lower costs per unit, this is **not** a case of agglomeration economies because the entrepreneur has control over the benefits because the outcomes are within the control of the firm. In contrast, agglomeration economies are present if three plants, each owned by separate parties, settle in the same location and receive benefits from these independent decisions.

C. External Economies of Scale (Internal to Industry, External to Firm)

External economies of scale that are internal to the industry but external to the firm occur when many firms in the same industry locate in an area to capture the external benefits of clustering. External economies of scale that are internal to industry are equivalently called **localization economies**. The scale of importance here is the size of the industry, rather than the size of each individual firm. For example, industry may concentrate in a locale to access a specialized labor force. The brewing industry that burgeoned in Milwaukee, WI is one historical example. Other benefits include waste processing, which may not be financially feasible on a small scale, research and development, access to raw material markets at a level that implies lower cost for all firms, and specialization in inputs (Nourse, 1968). The development of specialized inputs is another important factor. As industry expands at a site, it is possible that the production of vital components of production are provided by one firm rather than by each firm. This specialization offers the input at a lower cost to all businesses in the area.

Localization economies can also occur on the buyer side of production and tend to develop in output markets when the goods offered for sale are not perfect substitutes, variation between products is subtle and best assessed by inspection. In these cases, competitors may be drawn to locate next to one another despite their rivalry (Hoover and Giarratani, 1984). This phenomenon explains why furniture, car dealers, shoe stores and one-stop shopping malls agglomerate. Benefits accrue to firms because overall aggregate sales increase when consumers save travel time and are easily able to compare subtle differences in style, price and quality of goods.

Whether in the output or input market, the key element to localization or external economies is that they accrue for all the firms in a single industry at one location because of the increase in total output or sales of an industry (Carlino, 1982). It is not necessary that firms locate in an urban area to capture localization economies. So long as firms are clustered, the benefits of localization economies can be received even in isolated places.

D. External Economies of Scale (External to the Firm and Industry)

The broadest form of agglomeration is economies of scale external to both the firm and industry, also known as **urbanization economies**. These are the cost savings experienced by all firms when business activity is concentrated in metropolitan areas. As Carlino describes, urbanization economies come about "for all firms in all industries at a single location, consequent upon the enlargement of the total economic size (population, income, output or wealth) of that location" (1982, p. 97). Unlike localization economies, which grow as the number of firms located at a site in the same industry increase, urbanization economies are a function of ***city*** size. This is true because the benefits of a metropolitan location increase as the wealth of services rises, which is in turn directly proportional to city size (Hoover and Giarratani, p. 121, 1984). These benefits are not related to the size of the individual firm or the size of the industry cluster.

External benefits of urbanization come in many forms, and it is this fact that makes measurement of agglomeration benefits difficult. Economies of scale are achieved in a variety of ways. Firms are closer to markets, or at least closer to the transportation and communication apparatus necessary to shuttle their product to its final destination. These are transfer savings. Cities provide firms with greater access to financial services, capital, infrastructure and labor. Smaller inventories are needed with the proximity of wholesale sources. Specialization in production is possible on a wide scale. As each firm specializes and thus achieves economies of scale, the products and services produced are made available to all purchasers at a lower cost than without urbanization economies (Hoover and Giarratani, 1984).

Economies of scale also exist in the infrastructure needed to support business operations. Electrical power, water, roads, sanitation services and fire and police protection can all be provided at a lower cost per unit than would occur without urbanization (Isard, 1956).[5]

A few caveats about the concept of urbanization economies are needed. First, as Meyer (1977) emphasizes, once the firm is able to capture and control the external benefits from urbanization economies, those benefits of agglomeration are better considered as internal economies of scale. Second, the existence of strong agglomeration benefits brought on by urban location does not preclude the possibility of external dis-economies. For example, a firm may locate in a city and produce specialized inputs for an industry, which results in lower average costs of production for all firms using the input. But this same firm may emit pollution at a level that exceeds the marginal benefit it brings in the form of cost savings to firms.

Third, given the internalization of benefits and the existence of dis-economies, location patterns are a dynamic process in which firms shift to metropolitan areas to capture external economies while other firms, facing growing dis-economies, may make location decisions that move operations to the perimeter of the city and beyond. Finally, many external benefits of a city location may not be the result of urbanization economies but instead localization economies. Firms that locate in the city may be attempting to be close to other firms in the industry rather than attempting to capture the cities' amenities. In such circumstances, a mass movement of all firms in the industry to a less urban setting would not adversely affect the industry cost structure.

III. Alternative Explanations of Firm Location

Although there has been significant attention devoted to the role of agglomeration economies in firm location, the term has also become a convenient "repository for the untidy aspects of location theory" (Bopp and Gordon, 1977, p. 125). Vague aspects or less understood concepts of a firm's selection of a site are attributed to agglomeration. Several compelling alternatives that may play a significant role in firm location are presented.

A. Search Costs

Pascal and McCall (1980) contend that while agglomeration occurs as a response to the advantages of clustering, the theory of agglomeration economies cannot fully account for the location decisions in urban areas. The authors note that the availability of specialized labor is one strong argument for agglomeration but, given modern communication, it accounts only for why firms locate in the same general area and does not explain why so many firms locate in close proximity to one another. They also argue that the

[5]It is also true that these services could be obtained at a comparable low price if firms cluster in non-metropolitan areas, i.e. localization economies.

current distribution of firms is much heavier in concentration than would be expected if business is only receiving the direct cost savings of agglomeration economies alone.

Pascal and McCall turn to the importance of search costs in determining the location for a firm entering an industry. If perfect information is assumed, the firm knows the best place to locate. But since information is not perfect, new firms try to minimize their risk by observing and imitating the location decisions of successful businesses. The authors argue that location is driven more by the uncertainty of information that new firms have about their production function, supply network, market and technology. Imitating behavior is the dominant strategy, particularly for smaller businesses:

> Adrift in a sea of uncertainty, what could be more logical than heading toward those islands which have demonstrably provided salvation to similar and earlier swimmers? It costs so much for the new firm to acquire information about its optimal location that its managers will naturally be drawn to sites where rival firms appear to thrive (1980, p. 386).

The importance of search costs in the location decision merit further study, but as Pascal and McCall admit, search costs may explain why new firms locate next to existing business, but cannot address why these older establishments located at a particular site in the first place.

B. The Evolution of the Modern Firm and Its Economic Environment

A firm that locates or chooses to expand its operations in a metropolitan area may be responding to the inevitable evolution of industry rather than to agglomeration benefits. There are three fundamental changes in the U.S. economy that have implications for firm location (Luke *et. al.*, 1988)

i. *Movement from a goods-producing to a services-producing employment base*
ii. *Shift from a national to a global economy*
iii. *Transition from labor-intensive to technology-intensive manufacturing*

Location theory has always emphasized the geographical balance a business seeks between access to markets and access to raw materials. Because the products of a service firm are less tied to raw materials, this liberates firms from the traditional location "tug of war". Service industries are freer to locate near their markets. By the very nature of the service industry, this will be in areas where there is a concentration of population. Given the concentration of markets in highly-populated areas, the increased migration of firms to urban environments can be seen as part of this service orientation.

The increased focus on international trade and markets has stepped up the pressure on firms to develop strong communication and trading networks. Modest-sized business now exports products and may import a host of components for its own production. This internationalization of business for many firms increases the advantages of locating in or near "world cities," urban areas with links stretching across the country and the globe (Taylor, 1973). Thorngren (1970) suggested that modern communications could lead to a decentralization of business location as firms move out of high-rent urban areas to lower cost, small cities with comparable communication technology. But in many cases, face-to-face contact with other businesses and immediate access is crucial. These businesses, particularly if they are service oriented, utilize more complex information because products must travel through complicated trade routes.

The need for a wide array of modern communication services by many firms emphasizes the difference in the type and quality of information available in urban versus rural environments. Manufacturing operations, which tend to be more programmed and routinized in process may not need the face-to-face contact nor the wide array of information of the city and this is likely a strong reason why many facilities locate in other areas. But Westaway (1974) notes that there are innovative and creative aspects involved in a firm maintaining its market share and that these aspects often rely on the quality of information and ideas circulating in the city. Because service goods are not as standardized as the routinized goods produced on the assembly line in a manufacturing industry, the type of information required for operation is not trivial.[6] Firms operating in volatile and particularly competitive industries may absolutely require a city location to keep abreast and quickly respond to business developments. Westaway (1974) reports that a survey of prominent West London business found ease of contact was given as the most important variable in city location.

Many of the technology advances in industry have reduced the demand for semi-skilled workers. Skilled labor is becoming a critical input in production. The additional effort required to obtain advanced education dictates that trained labor is scarce, but more likely found in urban areas. This occurs because advanced training opportunities are usually found in larger cities, and job seekers are more likely to stay in a metro area because the wider array of companies increases the probability of finding work. This implies a firm will have better success in attracting the workers it needs if it is in or near metropolitan areas.[7] While it can be argued that access to specialized labor is simply an example of agglomeration economies, there is an important distinction. Agglomeration economies point to the additional ***advantages*** urban areas may have for providing specialized labor. This is distinctly different than the emerging ***requirement*** imposed by technology advances that firms choose a city location or forfeit access to the inputs it needs. The latter is imposed because of the transitions occurring in manufacturing, while the former is an external benefit accruing to the firm.

A final critical aspect of modern business is the trend of physically separating firm operations. Recognizing that various activities of the firm require different amenities, a division of systems tends to locate routinized activities in low rent regions and information-intense operations in urban areas. This is chiefly represented in the location of head offices in urban areas while repetitive tasks are relegated to more rural locations where fixed costs and semi-skilled labor costs are lower. Firm headquarters in urban areas are necessary to access the many specialists that support business such as lawyers, accountants, business consultants, advertising, banking and university research (Nourse, 1968). Westaway argues that this trend is merely a reflection of the emerging "hierarchy of corporate decision-making" (1974, p. 147). He continues:

---

[6]For example, manufacturing a simple nut or bolt is a programmed activity that involves little innovation. An out-of-city location could likely meet the low-level information needs of such a firm. This is contrasted with a graphic arts firm that designs logos for businesses. The product is not uniform as it is produced only after much face-to-face contact.

[7]For an example of this phenomenon, see Scott's (1992) case study of aircraft plants in Los Angeles.

> In short, the underlying cause. . . .has been the way in which business corporations have evolved. . . .. The main forces behind these tendencies appear to be the information requirements of different types of organizational units and the information richness of cities in different parts of the country (p. 148)[8]

C. Non-Economic Considerations

Theory tends to focus strictly on the economic factors that culminate in a firm's decision to locate in a particular area and agglomeration economies has emerged as a dominant explanation of firm location decisions. In practice, firm's location decision may not be rigorously scientific nor adhere to a recognition of the benefits of agglomeration economies. Blair and Premus (1987) stress that the location decision of firms cannot be viewed as an isolated activity but as part of an intricate, overall corporate strategy. For an existing firm, if shortage in capacity is anticipated, building an additional facility is only one of many options the business may select (Schmenner, 1979). If another site is determined to be necessary, the firm typically assembles a site selection team to look at a variety of non-economic and economic criteria. Even the decision of who serves on the team can affect the outcome of which site is selected (Blair and Premus, 1987). For example, one-third of all Fortune 500 companies in the U.S. rely on outside consultants to assist in location decisions, and the tools and techniques used by these agencies are diverse.[9]

Economic criteria include a firm's desire to enter new markets, corporate interest in integrating or segregating its activities and desire for company visibility. Non-economic factors are also important. In terms of headquarter location, one survey examining the location decisions for corporate headquarters found that CEOs have substantial opportunity to exert their personal preferences in the choice of a location (Blair and Premus, 1987). In a mail survey reviewed by Blair and Premus (1987), 41 percent of respondents indicated that the personal reasons of management were a critical consideration. For existing business, these concerns were an important influence for 29% of survey respondents.[10] New businesses tend to locate where the entrepreneur is from and not necessarily where the enterprise has the best chance to absorb agglomeration benefits. The role of these subjective criteria in location choices underlines the importance of not mistaking location decisions as purely stemming from agglomeration economies. Qualitative factors such as historical precedent, an area's business climate, education

---

[8]The relationship between location and the dispatching of firm activities to different locales has important implications for location behavior and has not been widely studied. However, it cannot explain the location decisions of, for example, small entrepreneurial firms that do not start their operations at a scale dictating a geographical separation of functions and locate in the urban area despite the higher costs of production there.

[9]One consulting firm gathers 22 indicators of industrial manufacturing climate (some of which may proxy agglomeration advantages). States are ranked based on how much their amenities match the needs of manufacturing associations. By this method, two of the highest-ranking states in 1985 were South and North Dakota. Based on their low population density and lack of significant urban centers, these states are not likely to offer the significant agglomeration advantages that economists claim motivates much of a firm's location decision.

[10]The mail surveys tended to focus on smaller firms, thus possibly slightly exaggerating the importance of personal considerations in the business decision-making process.

facilities, attitude of work force, cultural attributes of an area, government cooperation and commuting distances (Schmenner, 1979) may be the source of primary motivation, rather than the advantages of agglomeration economies, in the location decision. Krugman (1991) posits that non-economic factors create initial activity, and then cumulative causation gives rise to industrial concentration as firms and workers both prefer larger markets.

IV. Agglomeration Economies: A Survey of Empirical Work

Acknowledging the importance of forces beyond agglomeration economies assists in understanding factors firms consider in choosing a location. The existence of these forces has not, however, detracted from the emphasis placed on agglomeration theory in location choices. Much work has been devoted to trying to characterize agglomeration benefits in operation. In general, these efforts have been challenging; suspecting the importance of agglomeration economies in location decisions is different than measuring their existence. This section reviews industry surveys, case studies and models that attempt to **measure the existence** of agglomeration economies.

A. Surveys

Direct survey of firms is one way to measure the importance of agglomeration factors in encouraging firm site locations. This can be difficult to do because it is obviously not possible to ask firm management, "Did you locate at site x to capture the external benefits of agglomeration?" Proxies that are generally used to identify agglomeration forces include questions about neighboring suppliers of inputs, area labor availability and closeness to markets. In general, accurately constructing a survey to use as a data set for testing location hypotheses has been problematic (Dorf and Emerson, 1978).

In their review of survey efforts, Blair and Premus found that over half of all location possibilities that firms consider are eliminated at the state level. At this tier of consideration, the important factors tend to be labor, state tax policy, climate of region, and proximity to markets. In a review of a survey of location decisions of U.S. firms in the 1940s and 1950s the most important variables considered were access to markets, labor and raw materials. Other factors having little significance include taxation policy of the region and financial incentives (Blair and Premus, 1987). More recent work includes ***Fortune*** magazine's review of the location decisions of 1,000 of the largest U.S. industrial firms. Results revealed that in considering where to build a mainland branch plant, market access was considered as the most important factor. When firm management was asked to consider the factors that inspired previous plant locations, they identified labor availability as a chief consideration; an indication that markets, and hence, agglomeration economies may play a more important role in the future than they have in the past.

In a 1981 mail survey reviewed by Blair and Premus, new and existing firms were asked about their process of comparing various interstate sites. New firms were most forcefully swayed by the presence of strong markets, followed by labor force and transport concerns. Access to capital was also important. For existing firms expanding in another site, access to a low-cost and readily available labor supply was the most important concern, followed by market and raw material access. A 1984 ***Industry Week*** survey of 1,000 corporate managers found that the most predominant locational factor across a variety of businesses was transportation.

In general, the importance of factors in location decisions varies across firm type. High tech executives ranked as a low priority the transportation of inputs and market access but considered access to specialized labor and proximity to universities as the first and third most important criteria respectively (Blair and Premus, 1987).

In a forthcoming article, Barkley and McNamara analyzed the survey process for inconsistencies between actual location decisions and the locational preferences indicated in surveys. They found consistent information for easily observed criteria such as highway proximity. For other characteristics, they found consistency when the data were disaggregated by plant size and labor skill requirements; indicating resource constraints for small firms and special consideration of those factors that firms felt were of particular importance. They also found that while non-economic characteristics may not attract firms, they may dissuade firms from locating in a given community.

B. Indices

Indices have also been used to predict areas where agglomeration economies will be strongest. The most elementary approaches utilize concentration of industry as some measure of agglomeration. In a brief review of various index specifications, **Wheaton and Shishido (1981)** select a simple version to incorporate into an agglomeration model:

$$H=\sum_{i=1}^{n} \frac{P_i^2}{P^2} \qquad (1)$$

where:

P	=	Total number of firms
P_i	=	Number of firms in industry i
n	=	Total number of industries[11]

This specification can be used to measure agglomeration if it is assumed that concentration of industries is equivalent to industries being linked to one another.

Dieperink and Nijkamp (1988) developed a more sophisticated index to analyze the effects of agglomeration economies on innovation and research and development. The index, expressed on a scale of one to nine, is a function of three variables, city size, and distance to the main city center and inter-urban influences (distance to neighboring city centers) and is expressed:

[11]The authors present this index in a general sense where n is the total number of categories under study. Here, we identify the category as industries.

$$Agglomeration\ Index = f_I\left(\prod_{k=1}^{K} A_k + w * \prod_{j=1}^{J} B_j\right) \qquad (2)$$

where:

A_k	=	The distance to the kth closest (high population) city
B_j	=	Distance to jth closest (low population) city
w	=	Weighting factor that ranks the city for its relative importance
f_I	=	Index function that transforms values into integers

The authors did not find a significant relationship between innovation and their measure of agglomeration.

Marcus (1965) measures agglomeration benefits as the residual value of growth rates in a study of industry performance in New Jersey. The model assumes that the ratio of the growth rate of industry j in some region relative to the growth in population of the state will equal the ratio of national industry growth to national population growth:

$$\frac{G_j^{NJ}}{G_P^{NJ}} = \frac{G_j^{N}}{G_P^{N}} \qquad (3)$$

where:

NJ	=	New Jersey
P	=	Population
N	=	Nation
j	=	industry j

If the industry grows at the expected rate, then the difference between the growth rates will be zero. If this difference is non-zero, then this residual is assumed to be solely a function of agglomeration benefits (A_j):

$$G_j^{NJ} - G_P^{NJ}\left(\frac{G_j^{N}}{G_P^{N}}\right) = A_j \qquad (4)$$

This method is similar to an index in that results can be ranked according to whether they possess high or low agglomeration benefits. As with an index, the construction assumes rather than demonstrates that agglomeration forces are the chief source of residual differences in growth patterns.

More precise measurement of the existence of agglomeration economies examines the linkages between industries and is careful to distinguish between industrial concentration and relationships between those industries that may initiate agglomeration benefits. The central hypothesis is that if agglomeration economies are important in the firm's choice of location, then pairs of industries that are linked by flows of goods and services will tend to locate close together. No agglomeration benefits are assumed if linked firms do not choose to operate in the same area. To determine which industries tend to situate together, correlation coefficients or factor analysis are used. Results provide a gauge of **spatial association**. However, correlations express geographical tendencies rather than functional associations; consistent spatial associations do not imply that there are **economic linkages** between sectors.[12] The location tendencies of unrelated industries must be separated from those of strongly connected industries to decipher inter-related firms. This has most typically been achieved by using input-output technical coefficients.

Using factor analysis, **Bergsman *et. al.*, (1972)** analyzed clustering patterns in both non-manufacturing and manufacturing industries across the U.S. in 1963. This work marked an early effort to characterize the tendencies for linked industries to locate in particular areas. The authors found that eight cities were home to industries not found in other large metropolitan areas. Overall, however, it was not possible to unequivocally identify linkages between firms that were consistent over broad geographical areas. For example, it was found that the metal can industry tends to locate where food processing is dominant, but the industry also frequently locates in small cities that specialize in metalworking. The authors concluded, "Clustering is not mutually exclusive. Although many activities are clearly associated with one and only one cluster, many others are not" (p. 287).

Lever (1972) constructed a correlation matrix to analyze sixty-one industries in sixty-two U.K. subregions.[13] Findings indicated that industry location is not random and that marked clustering patterns are common. Using total sales data, an input-output matrix was developed to identify economic linkages.[14] Lever found that 24 percent of all industries had either mutual or one-way functional linkages. These connections were strongest for heavy, low-value goods and for the engineering industries. Because positive spatial and economic linkages support the existence of strong agglomeration economies, Lever noted that policy aimed at encouraging firm relocation to areas of unemployment might not be successful if the cost savings of clustering, particularly in transport, are ignored.

---

[12]Two industries or firms can be strongly correlated and have nothing to do with one another. For example, export shipping and the fisheries industry may be found together because both need to locate in coastal areas. But neither industry strongly relies on the other for inputs of production nor shares a common output market.

[13]Observations used to construct the correlation matrix were:

$X_{ij} = x_{ij}/x_i$

where:

x_{ij} = employment in industry j, region i

x_i = total manufacturing employment in region i.

[14]A matrix of industrial groups is created in which each element is:

$T = \Sigma t_{ij} + \Sigma t_{ji}$

where:

t_{ij} = total value of sales by firms in industry i to firms in industry j

t_{ji} = total value of sales by firms in industry j to firms in industry i

Functional linkages were defined to exist in pairs which conduct at least 1.65% of total trade with one another.

In an analysis of French and German industries, **Streit (1969)** developed a regression model to explain spatial association as a function of economic linkages and an agglomeration variable. Simple correlations of employment between industries were used as the dependent variable. A linkage variable, L_{ij}, was measured as follows:

$$L_{ij}=L_{ji}=\frac{1}{4}\left(\frac{O_{ij}}{\sum_i O_i}+\frac{O_{ij}}{\sum_j I_j}+\frac{O_{ji}}{\sum_j O_j}+\frac{O_{ji}}{\sum_i I_i}\right) \quad (5)$$

where:

O	=	Output, where i and j are industry linkages
I	=	Input
L_{ij}	=	Measure of linkage between industry i and industry j

An agglomeration variable is also developed to account for the fact that industries may not be closely linked but may move to an area to capture some of its positive locational attributes. It is defined as:

$$r_x=\frac{1}{2}(r_{kx}+r_{jx}) \quad (6)$$

where:

x	=	Industry as a whole in the region
k	=	The kth industry in the region
j	=	The jth industry in the region ($j \neq k$)

This variable (r_x) is a simple average of two industry correlations. The first measures how much industry k is associated with industry as a whole. The second measures j's proclivity to be found in areas where industry as a whole is found. Together they comprise a measurement of the affinity the two industries have for other industry. Incorporating this agglomeration variable into the regression, the model becomes:

$$r_{kj}= \beta_o+B_1L_{kj}+B_2r_x+\epsilon \quad (7)$$

where:

r_{kj}	=	spatial association between industry k and other industries, j=(1,...n)

This specification fit the data well for seven of 26 French and German industries. Particularly in French industry, the agglomeration variable proved to be the most significant contributor to spatial association.

Bopp and Gordon (1977) argued that Streit's model only captures direct linkages between two firms and that frequently location can be explained by the fact that industries cluster because they rely on a third party, buying from the same source or selling in the same market. Using 1967 U.S. industrial data, the authors constructed a total requirements input-output matrix (flows of goods to and from each industry). Their model is:

$$r_{ij}=B_o+B_1P_{ij}+B_2P_{ji}+B_3R_{ij}+B_4C_{ij}+B_5r_x+\epsilon \quad (8)$$

where:

r_{ij}	=	Correlation of industry i and industry j
P_{ij}	=	Input-output coefficient from industry i to j, weighted by transport costs
P_{ji}	=	Input-output coefficient from industry j to i, weighted by transport costs
R_{ij}	=	Correlations of paired row coefficients, (measures forward linkages)
C_{ij}	=	Correlations of paired column coefficients, (measures backward linkages)
r_x	=	Agglomeration variable, as defined in equation 6.

While the adjusted R^2 was low (.33) and *F*-statistic insignificant, the authors found the agglomeration variable to be positively significant but were not convinced that the agglomeration effect "explain(s) very much in the way of transaction or economic linkages." (p. 127).[15]

In another study using survey information, regression and factor analysis, **Dorf and Emerson (1978)** examined the criteria that manufacturing firms use for selecting non-metropolitan plant sites in seven mid-west states. Their results stress the value that firms place on amenities that are positively related to city size. While production facilities did locate in less inhabited areas, in choosing between similar sites, firms considered the size of the community, its distance from urban centers and the availability of labor to be the most important factors.

Models that explain the movement of goods and services have been developed in the field of geography. The best known approach is the gravity model. Gravity models assert that there is an inverse relationship between movement and distance **(Zipf, 1949)**. Interaction and trade can be explained by the length and intensity of movement required, and firms will try to minimize their effort expended. Taylor (1973) adapted this idea to analyze agglomeration economies for iron foundries in two regions of the U.K. The author asserts that trading patterns are a function of eight interacting variables:

- Size of supply source or production center
- Size of market for the good
- Competition between suppliers
- Distance between consumer and supplier

[15]Of course, both correlation coefficients and input-output analysis have drawbacks in location applications. Correlation coefficients are highly sensitive to the level of disaggregation in the data. In addition, correlations can only describe paired associations while it is the actual clustering behavior of industries that is usually of interest. Both input-output and correlation analysis use data that defines industries via geographical boundaries (e.g. counties or states) when the relevant unit of observation is a functional economic unit which may not be neatly described by these divisions. For a review of input-output's other shortcomings, see Hamilton et al, 1991; Nijkamp, 1987, p. 317.

- Product variability (in type or quality)
- Spatial variation in price of product
- Ownership factors that may predetermine linkages
- Agglomeration factors

Taylor's model is:[16]

$$C_{ij}=\frac{\frac{s_j}{d_{ij}}}{\sum_{j=1}^{n}\frac{s_j}{d_{ij}}}*N_i \qquad (9)$$

where:

C_{ij}	=	The number of consumers in location "i" likely to purchase good from production center "j"
s_j	=	Size of supply source (indexed)
d_{ij}	=	Distance from consumer at "i" to production site "j"
N_i	=	Number of potential consumers (indexed)

The specification of the model penalizes a supplier that is far away but rewards it if its supply is significant. The author explicitly incorporates distance and the size of the supply source. By summing supply across all firms, competition is implicitly incorporated. Because data is lacking for price variations, ownership and variation in product type, the author did not incorporate these variables into the model. Distance is proxied using transportation costs, and size of supply and the number of consumers are indexed using census data. By excluding any measurement of agglomeration, Taylor simulates the expected trading patterns that would be seen in the iron foundry industry if no external economies of location exist. He compares the expected results from the model for both iron foundry sales and purchases against the trading patterns reported by surveying approximately 55 percent of the iron foundry industries in two regions. These industries were asked to identify their five main customers and five chief suppliers of materials. Volume of trade was not considered.

By plotting the residuals of expected and observed trading patterns of iron casting the author found pronounced differences that he attributes to "the immediately local industrial environment" (p. 393). In the West Midlands, the model predicted that 25 percent of all sales would be local, while survey results indicated 42 percent of the sales linkages between iron founders were local. In the second region, Lancashire, agglomeration economies were stronger, with the model predicting that 11 percent of all sales would occur in the nearby vicinity of the industry while the survey results showed that 62 percent of all sales took place within a close perimeter. Taylor comments, "external economies. . . .are of sufficient strength to cause its iron founders to be six times more strongly linked with juxtaposed customers than

[16]The author also includes a parameter which measures consumer's willingness to travel to acquire different types of goods. Because the analysis sets this parameter equal to 1, it is not included in the model's specification.

the model would anticipate." (p. 393). The most concentrated linkages were found in the engineering industries.

Given the myriad of inputs for iron casting, difficulties arose in accurately measuring the backward linkages that exist in the industry. The author obtained data on three major resources, pig iron, scrap metal and molding sand equipment. Tendencies for local linkages spurred by agglomeration were less palpable for purchasing of inputs to iron foundries. There were no distinct differences in the observed and expected linkage patterns for pig iron and molding sand supplies. Only slight agglomeration tendencies for scrap iron were seen.

V. Measurement of Agglomeration Economies

The identification of firm or industry clusters serves as a basis for concluding that agglomeration benefits exist. The studies addressed above generally attempt to measure agglomeration through the measurement of economic linkages. This approach can be challenged. Meyer (1977) warns that linkages may or may not be a good measurement of potential agglomeration benefits. He writes, "Some scholars imply that linkage within a location (city or region) is *prima facie* evidence of the existence of external economies" but linked growth of industries may be due to factors having nothing to do with external economies (p. 85).

More rigorous work has tried to directly measure agglomeration economies. The following section reviews the models and techniques that have been developed to quantify the external benefits of localization and urbanization. Generally, these models are divided into three categories, production functions, other econometric models and theoretical models.

A. Production Function Models

There are three broad approaches used in the measurement of agglomeration economies through production functions (Duffy, 1987):

i. Incorporating variables that proxy city size in a Hicks-neutral shift parameter for a homogenous production function applied to cross-sectional urban data. **Agglomeration is measured in the constant term of the production function.** (*e.g.*, Segal, 1976; Sveikauskas, 1975; Moomaw, 1985; Nakamura, 1985; Greytak and Blackley, 1985).

ii. Estimating returns to scale for small versus large cities using a variety of homogenous production function specifications. **Agglomeration is measured in differences in the returns to scale parameters**. (Shefer, 1973; Carlino, 1979; 1982).

iii. Variations on i and ii including:

- Fitting non-homogenous specifications (Duffy, 1987)

- Using factor-augmenting specifications (Calem and Carlino, 1991)
- Measuring production functions using state-level data (Beeson, 1987)

The skeletal model that is used to measure agglomeration economies through production functions is:

$$Q=g(\cdot)f(K,L) \qquad (10)$$

where:

Q	=	Output
$g(\cdot)$	=	Scale parameter
K	=	Capital
L	=	Labor

As per the approach in i., each city or geographic area is assumed to have its own aggregate production function, differing from other cities only by the scale parameter. This is an important point. Analysis that measures differences in the scale parameter assumes that the functional form of equation 11 is identical for all areas. That is, technology and inputs are assumed to be constant across space. There exists a family of uniform production functions, each differing in the level of the parameter. Possible differences in the marginal products of labor and capital are not captured using this approach.[17]

Specifically, the constant term, typically defined in neoclassical economics as the technology parameter, is defined as a scale effect that describes the increases in output that accrue because of an urban location. Differences in production across cities stem from differences in this scale effect, which is measured through the use of a surrogate variable. The literature has tended to use population as a proxy for this scale effect. In this case, then, the parameter is notated g(n) where n=number of people living in an urban area. Implicitly, the first partial is assumed to be positive - $g'(n)>0$ - implying that there is some efficiency associated with a city size, as measured by population. This assumption has been the subject of keen debate in the literature.

A further important assumption generally made about the scale effect is that it exhibits Hicks-neutral productivity *i.e.*, a technical change affects all factors of production equally. For a Cobb-Douglas production function this implies:

$$Y=e^{\lambda}(x_1^{\alpha}x_2^{1-\alpha}) \qquad 0<\alpha<1, \qquad \lambda>1 \qquad (11)$$

The exponential term, λ, expresses some level of technical change. Note that the impact of the change on the production function does not alter the input ratio. Average factor productivity increases but the marginal rate of substitution between inputs is unchanged.

[17]The validity of this assumption depends largely on whether innovation is equally available in all locations.

This is contrasted with a factor-augmenting technical change which exclusively effects the efficiency of one input. Most attempts to fit production functions to output data for cities assume Hicks-neutrality. This has been questioned by some who argue that factor bias may be more likely (Carlino, 1979).

A second approach (ii) focuses on the returns to scale as a measure of agglomeration levels.[18] The literature has emphasized that returns to scale are a precise measure of agglomeration economies. If agglomeration is beneficial, then higher returns of scale should be found for larger cities. In the words of Carlino:

> If, as hypothesized, agglomeration economies are largely technical in nature, due to increasing returns, then the homogeneity parameter of the production function serves as a direct measure of this phenomenon (1989, p. 366).

For both approaches, f (equation 11) is generally defined as a homogenous function of degree 1 such that returns to scale are the same at all points along the expansion path.

Much of the estimation of agglomeration economies begins with the CES production function. The general form of the CES model is:

$$Q=A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\frac{\theta}{\rho}} \tag{12}$$

where:

Q	=	Output
L	=	Labor
A	=	Efficiency parameter
δ	=	Share of income going to each factor ($0 \leq \delta \leq 1$)
ρ	=	Substitution parameter
θ	=	Degree of homogeneity parameter

Returns to scale can be derived by multiplying K and L by a constant, λ:

$$\begin{aligned} Q &= A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\frac{\theta}{\rho}} \\ & A[\delta(\lambda K)^{-\rho} + (1-\delta)(\lambda L)^{-\rho}]^{-\frac{\theta}{\rho}} \\ &=A\{ \lambda^{-\rho}[\delta K^{-\rho} + (1-\delta)L^{-\rho}]\}^{-\frac{\theta}{\rho}} \\ &=A\lambda^{\theta}[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\frac{\theta}{\rho}} \end{aligned} \tag{13}$$

[18]Occasionally, attention is also given to the elasticity of substitution between factors for a given change in prices (assuming firms are cost minimizers).

Since the bracketed term in the last line of the equation multiplied by A is equal to Q, returns to scale are:

$$\lambda^{\theta} Q \tag{14}$$

If:

$\theta = 1$ constant returns to scale
$\theta > 1$ increasing returns to scale
$\theta < 1$ decreasing returns to scale

Depending on whether approach i. or ii., is adopted, attention is paid to either specifying a functional form that describes how the efficiency parameter changes across spatial areas or measuring the returns to scale across cities for the same industries. For either emphasis, a specific function form must be assumed for equation 15. Frequently the Cobb-Douglas model, a specialized case of the general CES model, is used.

Another important specification is the Dhrymes formulation, which is a one-factor labor model that allows estimation without capital data. Dhrymes shows that if perfect competition is assumed in the input market, the CES model collapses to:[19]

$$\ln(L) = \frac{-\rho}{\theta(1+\rho)} \ln(A) - \frac{1}{(1+\rho)} \ln(w) + \frac{\theta+\rho}{\theta(1+\rho)} \ln(Q) \tag{15}$$

In addition to the advantages of bypassing the need for capital data, Dhrymes is log-linear and can be easily estimated.

B. Limitations Using Production Functions

The use of production functions to measure agglomeration economies has been by far the most popular empirical tool. Generally these models have been successful in measuring the agglomeration economies accruing to clustering firms or to cities. Interpreting these results requires an understanding of some of the more challenging empirical problems facing these modelling exercises. A summary of the difficulties in the application of production functions is presented below.

[19]The derivation of the Dhrymes formulation is found in the appendix.

i. Data Problems

Most studies lack firm-level data and resort to using the industry as the unit of observation, employing two-digit SIC data, a relatively coarse level of classification.[20] This can introduce well-recognized problems of aggregation and the tendency for upward bias in returns to scale estimates (Peterson, 1990). As Hoover and Giarratani note, ". . . .many of the most interesting aspects of agglomeration economies can be appreciated only at a much more micro level." (1984, p. 125). Finally, nearly all work has been restricted to analyzing manufacturing agglomeration despite the growing significance of service industries in the economy.

The use of an average or representative firm as a way around the lack of firm-level data is also problematic because it inadvertently incorporates an industry size effect, precisely the external benefit that it is hoped will be captured in the estimation (Greytak and Blackley, 1985). As discussed further in the next section, estimates of capital are not always available or acceptable. The use of one-factor models, such as Dhrymes model, may introduce bias due to omitted variables. Significant problems of heteroskedasticity and multicollinearity, particularly between capital and labor and the various variables specified to represent city size, have been noted. Few contributions effectively deal with these estimation problems.

ii. Measurement of Capital

The problems inherent in using capital data as a measurable quantity begin with the general lack of agreement on exactly what constitutes a unit of capital. While capital is recognized as an essential element of the production process, it is only mentioned in general terms. In this paper we refer to capital as physical components regarded as produced means of production. The major problems pertain to capital valuation and the heterogeneous nature of capital.

Since capital goods can not be summed directly, researchers are forced to use capital value as an estimate. This value can be measured either in terms of its resale value, i.e. investment less depreciation, or in terms of the value of goods produced from the capital. Due to additional problems associated with the latter method, the former is the generally accepted approach. This method, however, leads to the problem of a changing capital value without a corresponding change in the capital stock, or vice versa. In either case, analyzing the effects of a given capital stock may be biased through the value placed upon it.

The heterogeneous nature of capital plays an increasingly problematic role as an analysis is conducted at highly aggregative data levels. This characteristic leads to the implication that two distinct forms of capital with similar values have similar productive capacities. There does not seem to be any a priori reason for this assumption. This problem can be abated by performing an analysis at a sufficiently specific level of detail. Unfortunately, the problem of data availability becomes an increasingly important issue.

---

[20]Only one article reviewed here, Greytak and Blackley use data that uses the firm as the unit of observation.

iii. Time-Series vs. Cross-Sectional Analysis

The literature contains a debate as to whether a time-series or cross-sectional methodology is more appropriate in efficiently estimating agglomeration economies. Both approaches have advantages and disadvantages while results derived from either method are highly dependent on the nature of their respective underlying assumptions. Time-series analysis is preferred by some researchers because it eliminates the possible effects of city variation. This approach also incorporates possible effects of industry composition as Moomaw suggests, "If industry composition is more stable over time than space, [time-series analysis] reduces the problems of changing industry mix." (1983, p. 527) Beeson (1987) employs a time-series approach to measure the effects of agglomeration economies on the growth rate of productivity. The major disadvantage to this method concerns the effects of technological change on the production process. How to best handle the dynamics of technology within the context of production functions is a debatable issue. Incorporating a technology parameter is possible, but interpreting how much of the constant estimated is due to change in technology and how much is due to agglomeration becomes difficult.

Cross-sectional analysis has the advantage of controlling for technological change. This approach takes technology as an exogenous variable while comparing cities or metropolitan areas to measure agglomeration economies. The disadvantages lie in the ramifications of the city and technology assumptions. Factors such as the production function and factor prices, as well as technology are assumed to be identical across cities, irrespective of size. Empirical support for these assumptions is minimal. In fact, Beeson (1987) has derived results that contradict the assumption of uniform technology. Her research supports the theories of both hierarchical and radial diffusion of technology. Together with the need for further evidence to either support or discount the assumption of city size as exogenous to the production function, analytic results must be interpreted within the context of these limitations.

iv. Question of Hicks-Neutrality

Homogeneous production function analysis assumes that Hicks-neutrality best describes the innovation process. Implications of this assumption include constant marginal rates of substitution among factor inputs. One argument that has been made in the literature is that this approach may obscure possible effects of capital intensive production. Capital-to-labor ratios may increase with city size. Consequently, agglomeration estimates may be biased. Another ramification of this approach stems from the theory of perfect competition and constant returns to scale. This approach has been widely used, but some authors, such as Greytak and Blackley, have challenged this assertion. They argue, "neutrality [is] an issue to be addressed empirically rather than an *a priori* condition of production." (1985, p. 1122). Duffy (1987) drops this assumption and analyzes nonhomogeneous production functions. His results indicate a sinusoidal relationship between returns to scale and city size.

V. Measuring the Components of the Scale Effect

There is significant debate in the literature over how to craft the scale effect. Two strains of argument can be perceived. The first, articulated by Carlino (1979), is that population is simply not a good proxy for agglomeration. Carlino argues, "[A]gglomeration economies refer to the external economies associated with size, not necessarily population size, and concentration" (p. 365). Others have disputed this concern (Moomaw, 1983).

The second concern largely accepts that population provides a reasonable surrogate for agglomeration but argues that modelling agglomeration through the specification of a scale effect provides only a gross estimate of the total impact cities have on production. Is the measurement a gauge of localization or urbanization economies? For example, if there is a significant difference in the scale effect between a city and a small town, how much is due to the benefits of the city (urbanization economies) and how much is due to industry agglomeration (localization economies). In addition, how can dis-economies be measured? If the differences in the scale effect for a small and medium city are small, could it be the case that agglomeration benefits in the latter ***are*** large but are offset by large dis-economies?

Although seemingly a small point of contention, these distinctions have important policy implications. Finding significant localization economies implies that there are benefits external to the firms but internal to industry. If these localization economies comprise most of the economies of scale a firm experiences, then relocation, especially of inter-linked clusters of industry, would not result in sizable increases in long-run production costs to the firms because it is the interdependence of industry, rather than city amenities, that is important. If most of the agglomeration benefits are due to urbanization economies - benefits external to the firm and industry - relocation will increase long-run average costs for firms and industry. Some of the more promising research, including Moomaw (1986), Nakamura (1985), and Begovic (1992) has tried to re-specify the scale effect to distill its aggregate approximation of benefits into three distinct effects: localization, urbanization and dis-economies. One suggestion is to adjust for industry size in order to make a distinction between localization and urbanization economies. When this is done, a further question arises in considering whether to use industry value added data to measure industry output or industry sales.

C. Review of Production Function/Agglomeration Literature

Keeping these shortcomings and debates over specification in mind, the following section reviews the empirical efforts to measure agglomeration economies through production functions. This review is done according to methodology. The discussion begins with those studies that use the Cobb-Douglas production function, then moves on to the application of the general CES production function. Some of the studies used the Dhrymes specification of the CES model so they are included accordingly. This section then turns to a discussion of studies that employ other econometric techniques. Table 1 displays a summary of the literature in a concise manner. Generally speaking, the analyses used 2-digit SIC data and/or SMSA population data in either a cross-sectional or time series fashion with exceptions noted below.

The Cobb-Douglas production function and its variations are preferred by authors such as Segal (1976), Kawashima (1977), Moomaw (1981, 1985, 1986), Tabuchi (1985), Greytak and Blackley (1985), and Nakamura (1985). Segal, Moomaw, and Greytak and Blackley examined agglomeration economies through labor productivity and factor returns. Moomaw analyzed 2-digit SIC data and employed comparative statics to determine that changes in productivity and wages affect the profit maximizing size of industries. His conclusions suggest that from 1967 to 1977, labor productivity declined in six industries and Hicks-Neutral productivity declined in four industries. Segal found CRS for SMSAs in 1967 while Greytak and Blackley found IRS for Cincinnati firms in 1975-76; however, they also tested the Dhrymes specification and found CRS. Segal's results suggest SMSAs larger than 2 million people benefit from factor returns eight percent higher than for smaller SMSAs. Moomaw includes initial capital stock in his analysis and reports increased returns of only six percent. Greytak and Blackley also find that labor productivity is higher and that firms operate with greater efficiency in large cities. Tabuchi used 1980 2-digit manufacturing data for Japanese cities to estimate urbanization economies; the key variable being population density rather than population count. He first considered urbanization economies being related only to capital and estimated an effect of population density on labor productivity of 4.3%. When including labor supply, the estimate increased to 8% for Hicks-neutral technology.

Authors who directly employed either the basic CES production function or a minor variation of it, aside from the Dhrymes method, include Shefer (1973), Sveikauskas (1975), and Calem and Carlino (1991). The returns to scale parameter is examined by Shefer who attempts to distinguish between localization and urbanization economies while Calem and Carlino make no distinction. Shefer employs both the CES and Dhrymes specifications with industry level data to estimate localization economies and then aggregates the data to estimate urbanization economies. Sveikauskas looks to population as the cause of agglomeration economies in terms of increased labor productivity. Shefer's conclusions suggest that urbanization economies exist for the entire manufacturing industry at an approximate rate of 20 percent while localization economies do exist but with a high degree of variation among the industries studied. Sveikauskas finds that a doubling of city size tends to increase labor productivity by approximately 6 percent, on average. However, Moomaw (1981) adds a capital-intensity measure to Sveikauskas' equations and derives an increase of only 2.5%. Calem and Carlino find that agglomeration economies exist for cities of less than 2 million people. They also find that technical progress is greatest in the largest cities and that the rate of technical change has declined since 1973.

Moomaw (1986) and Nakamura (1985) estimate the distinct effects of localization and urbanization economies while Kawashima (1977) considers only urbanization economies. Using data from 1958, 1963, and 1967, Kawashima estimates the optimal city size, in terms of benefiting from urbanization economies, is 5.5 million people. Moomaw finds that from 1967 to 1977, localization economies fell in five industries while increasing in only one; he also found that urbanization economies rose in three and declined in two industries. Nakamura, who had included land as a composite economic characteristic into the theoretical approach to his study using 1979 data, found that light industries tend to benefit more from urbanization economies while heavy industry tends to benefit more from localization economies. Nakamura attempted to infer an estimate of the land parameter which he found to be significant for one-half of the industries studied.

Hansen looked at the relationship between agglomeration economies and industrial decentralization. A survey performed in 1980 resulted in data for 356 manufacturing plants. His results suggest that a 1% increase in distance from the CBD corresponds to a .13% decrease in Hicks-neutral productivity. His work also suggests the observed decentralization is partially due to wage benefits that approximately offset the agglomeration economies of the central city.

Duffy (1987) compared the Cobb-Douglas form to two non-homogeneous production functions, the Ringstad model, which has unitary EOS, and the Vinod model which has nearly constant EOS. The specification for the Ringstad model is as follows:

$$\ln(Q) + c_0 Q + c_1[\ln(Q)]^2 = d_0 + d_1 \ln(\frac{K}{L}) + d_2 \ln(L)$$
$$RTS = \frac{d_2}{1 + c_0 Q + 2c_1 \ln(Q)} \tag{16}$$

The Vinod model and RTS are:

$$\ln(Q) = b_0 + b_1 \ln(K) + b_2 \ln(L) + b_3 \ln(K)\ln(L)$$
$$RTS = b_1 + b_2 + b_3[\ln(K) + \ln(L)] \tag{17}$$

Duffy uses capital, labor and output data for 49 of the largest SMSAs in the U.S. for 1975-77. Results indicate that the Cobb-Douglas specification is not as strong a model for measuring agglomeration economies as are the two other specifications. The most interesting finding is that returns to scale are not constant for either model and instead follow a sinusoidal pattern. Duffy suggests that relative RTS may be a more important factor than absolute RTS.

Carlino (1979, 1982) begins the Dhrymes specification of the CES production function in a time-series analysis. From this process he derives an estimate of the homogeneity parameter which he uses as a dependent variable to evaluate the effectiveness of population as a surrogate for agglomeration economies and in an attempt to isolate the effects of urbanization and localization economies. His results do not support the conclusion that population serves as a reasonable proxy for agglomeration economies, at least as measured by IRS. He then uses average manufacturing employment, the ratio of manufacturing employment in industry i relative to national manufacturing employment, the total number of reporting units in the jth SMSA, and population scale as estimators for internal scale, localization, and urbanization economies, and urbanization dis-economies, respectively. He concludes that localization economies are positive for five and negative for three industries; urbanization economies are positive for 12 and negative for 10 industries. In his 1982 study, Carlino repeated the process as before but used data at the SMSA level to analyze urbanization economies. His results for this study suggest that population is a reasonable proxy for urbanization economies. He also found an optimal city size of approximately 3.3 million people and a 7 percent decline in the optimal city size in the 1970's. The author suggests that the decline in optimal city size may be due, in part, to advances in transport, communication, and production technologies.

The final two production function studies are unique in their approach. Moomaw (1983) tests the theory that production functions vary at the industry level, while Beeson (1987) uses state level data to evaluate the relationship between agglomeration economies and productivity growth. Moomaw utilizes three different functional forms: (1) constant elasticity of productivity, (2) a single shift in the elasticity, and (3) declining elasticity. The differences among these three models lie in the form of the population variable used. For the constant elasticity of population, the logarithm of SMSA population is used. By including a dummy variable for an SMSA of greater than one million, the model allows for a shift in elasticity. The third formulation uses the reciprocal of population so that as population increases the elasticity of productivity decreases. His results do suggest some differentiation among industries. The constant elasticity form is best for one-half of the industries while the other two forms are each best for one-quarter of the industries examined.

Beeson employs a two-stage estimation procedure where the first stage consists developing growth rates for scale economies, technological change, and total factor productivity. Output growth is divided into CRS contributions of input growth and scale economies, with the residual being referred to as technological change. Total factor productivity growth is the combination of scale economies and technological change, or output growth less growth due to CRS. In the second stage she uses each rate as a dependent variable to isolate the determinants of each. She found that scale economies and technological change affect agglomeration economies and that the decline of manufacturing growth is due to a productivity advantage decline. Another aspect to this study is her inclusion of the spatial arrangement of cities. Beeson develops an index that incorporates the mileage among cities. This notion will be addressed further in studies concerning urban land use theory.

Carlton (1979) develops an econometric model to examine the location factors considered when branch plants are built or new firms are started. The analysis examines three manufacturing industries: plastics, electronic transmitting equipment and electronic components. He finds strong support for the existence of agglomeration economies. Due to the inherent differences between new firms and new branches, he develops two separate models. Carlton uses this model to examine births of firms in SMSAs for two periods, 1967-1971 and 1972-1975. The findings indicate that the agglomeration variable is positive, statistically significant and sizable in magnitude for all three industries but that it is "definitely not true that SMSAs with the greatest number of births are those with the greatest amount of existing activity" (p. 28). The most important factor influencing new firms in an SMSA is wages. A one percent decrease in wages increased new firms by one percent.

Begovic (1992) begins with the theoretical arguments for a unique urbanization economies function for a given city size, but industry-specific localization economies functions. Within this context, he uses 1981 data for 68 cities in an attempt to quantify the industrial diversification-city size relationship, as well as the industrial share-city size relationship. Results indicate a statistically significant relationship between industrial diversification and city size as well as industrial share and city size (except for transportation, tourism and leisure, and technical services). Results suggest growth of city size leads to an increase in urbanization economies. Also, finding indicate growth in city size leads to an increase in the localization economies of primarily service industries and a decline in the share of manufacturing (a decrease in net localization economies).

Wheaton and Shishido (1981) examine the question, at the international level, of whether there are economic causes for differences in concentrations in cities. They base their model on two assumptions: (1) the total size of a national market divided by the efficient market size of a city should equal the number of cities, and (2) the efficient output of a city depends on the level of economic development. Their data consists of 38 countries that have at least three metropolitan areas; due to data restrictions, estimations that include central government expenditures share are limited to 31 countries. The non-linear version performed much better than the linear version in terms of significance and consistency. Findings indicated up to the point of $2,000 per capita income, city development results in greater scale and agglomeration economies. Beyond this level, increases are still seen, but at a diminishing rate. Optimal population size for a metropolitan area was approximately 20 million.

TABLE 1
Summary of Empirical Agglomeration Literature

Production Function		
Authors	**Method**	**Findings**
Shefer 1973	Dhrymes and CES production functions. Homogeneity parameter varies.	Urbanization economies of 20% for manufacturing industries and varying localization economies.
Sveikauskas 1975	CES production function with CRS Estimates g(pop).	A doubling of city size increases labor productivity by 6%.
Calem and Carlino 1991	Estimates a CES production function with a varying homogeneity parameter.	Agglomeration economies for cities of less than 2 mil. and technical progress is greatest in largest cities.
Segal 1976	Cobb-Douglas production function	CRS across city sizes and SMSAs larger than 2 mil. had factor returns 8% greater.
Kawashima 1977	Cobb-Douglas production function variant (labor oriented)	Urbanization economies until city size of 5.5 million.
Moomaw 1981	Incorporates capital intensity and initial capital stock into the research done by Sveikauskas and Segal, respectively.	Adjusts the results of Sveikauskas and Segal to 2.5% and 6%, respectively.
Tabuchi 1985	Generalizes from Moomaw (1981) employing CES and Cobb-Douglas with population density.	Doubling population density will increase productivity between 4.3% and 8%, on average.
Moomaw 1985	Comparative statics and Cobb-Douglas production function to estimate productivity	The elasticity of labor productivity and Hicks-Neutral productivity declined in the 1970s.
Moomaw 1986	Cobb-Douglas value added production function	Mixed results for both localization and urbanization economies.
Greytak and Blackley 1985	Cobb-Douglas and Dhrymes production functions to compare alternative forms of capital data.	Labor productivity is higher and firms operate more efficiently in large cities.

Nakamura 1985	Cobb-Douglas production function variant (Translog) Includes land in the model development and then treats it as an omitted variable.	Light industries benefit more form urbanization economies while heavy industry benefits more from localization economies.
Hansen 1989	Cobb-Douglas production function	A 1% increase in distance from the city center corresponds to a .13% decline in Hicks-Neutral productivity.
Duffy 1987	non-homogeneous production function with constant EOS and Cobb-Douglas production function	RTS were found to be sinusoidal in nature. Relative RTS may be more important than absolute RTS.
Carlino 1979	Dhrymes production function for time-series to get an homogeneity estimate and then use it in a cross-sectional as the dependent variable with population.	Mixed results for localization and urbanization economies. Population is a poor proxy for agglomeration as measured by IRS.
Carlino 1982	Same as 1979 study	Optimal city size is 3.3 mil. There was a 7% decline in the optimal city size in the 1970's. Population is a good proxy for urbanization economies.
Moomaw 1983	Tests three different functional forms to estimate productivity: constant elasticity, a shift in the elasticity, and declining elasticity.	The constant elasticity form is best for one half of the industries, while the other two forms each fit one quarter of the industries best.
Beeson 1987	Developed growth rates for scale economies, technological change, and total factor productivity; then used them as dependent variables.	Found hierarchical and radial diffusion of technology. Scale economies and Technological change affect agglomeration economies.
Other Econometric		
Carlton 1979	Conditional logit, stochastic specification, and maximum likelihood to evaluate the affects of economic activity on the births of new businesses and branch plants.	Economic activity is positively related to agglomeration economies.

Begovic 1992	Examines the relationship between city size and both industrial diversification and industrial share.	Urbanization economies exist, localization economies in the service industry, and localization dis-economies in manufacturing.
Wheaton and Shishido 1981	Uses a non-linear algorithm and linear regression to estimate production localization functions.	Non-linear form is superior. Scale and urbanization economies exist up to per capita income of $2,000.

VI. Theoretical Models

The literature reviewed in this section does not attempt to measure agglomeration economies; no empirical tests are performed. Much of this work employs agglomeration as an input for explaining other phenomena. For the sake of simplicity, we refer to these models as theoretical. This terminology does not imply that the previously described work has no theoretical foundation. Rather, it is used to describe the whole of the general approach of this body of work.

Helsley and Strange (1991), in an attempt to isolate the underpinnings of agglomeration, analyze the role of risk in capital markets as a source of positive agglomeration economies. They contend that investors face risk in providing credit to investment projects and that the probability of an adverse state of nature drives the owner of capital to seek next best options for investment. Agglomeration economies exist because firms minimize the risk associated with investment and used assets are more productively employed because cities offer a variety of alternative uses and buyers of assets.[21] The model considers a single risk-neutral bank which chooses among alternative investment projects. Using this framework, the authors demonstrate that the expected value of a project is a function of entrepreneurial skill and rises with city size. They also derive the conclusion that the second best use of the asset, assumed immobile, is more productive in larger cities because of the variety of economic activities located there.

Urban land use theory, bridging location theory and urban economics, utilizes agglomeration economies to explain spatial land patterns. The work done by Fujita (1985a, 1985b), Ogawa and Fujita (1980), Papageorgiou and Thisse (1985), and Grimaud (1989) assumes agglomeration economies are elemental in describing firm behavior. Households and businesses compete for land so as to maximize utility and profits, respectively. The spatial interdependence of firms and households provides the basis for examining the nature and determinants of the spatial structure of cities. We focus here on research allowing for "spatial externalities".

Cities begin with an undeveloped tract of agricultural land on which households and firms bid for locations. Simplifying assumptions are linear cities, no relocation costs, perfect markets, and homogeneous space. Business firms are assumed to employ a homogeneous production technology depending on inter-firm communications, with communication costs being proportional to distance

[21]Note similarity between this argument and labor market theories presented earlier.

between firms. Productivity is therefore positively related to an individual firm's proximity to other firms. Labor is provided by households within the city. Agglomeration economies are incorporated into the model as a determinant of firm behavior through both the household and firm density distributions.

The typical analysis solves the following functions: household density, firm density, land rent, wage, commuting pattern, utility, and profit. The solution leads to the notion of equilibrium and optimal land use patterns. A significant result is that equilibrium city size is generally larger than optimal city size. The difference between the equilibrium and optimal conditions is revealed within the model through the land rent functions of the firms. When deriving the profit function for an individual firm, the degree of spatial accessibility is included as a positive factor for that firm. However, when considering the entire city, the notion of accessibility contains a duality in that each of two firms benefits from their mutual proximity. This feature is not included within the equilibrium calculations but is included within the optimal city size derivations. This difference or "location subsidy" is a consequence of spatial agglomeration. Thus, agglomeration economies are seen to have a direct impact on urban land use patterns and are vital in explaining urban spatial structure.

These theoretical analyses incorporate a type of localization economies (there is no distinction of industry) in a way that is distinct from either the production function or index approaches -- through transaction costs directly related to distance among firms. These relationships transcend the scope of any individual industry and therefore do not represent the traditional localization economies. Urbanization economies are included in a more subtle fashion and are not explicitly analyzed. Firms are assumed to desire a location near sufficient labor but the costs associated with locating near residential areas are included only on the household side, via commuting costs. A disadvantage of these theoretical analyses is that so far there has been little empirical testing of the hypotheses generated with the theory.

VII. The Role of Dis-Economies in Agglomeration

The analysis of agglomeration economies through production functions and other approaches supports the general conclusion that there are benefits that accrue to firms that locate together in urban areas. Observed migration patterns and the successful measurement of increase efficiency in urban areas seem to provide *prima facia* evidence that the pull of agglomeration is much stronger than the downside of congestion and other disadvantages attendant to urban life. The conclusion often drawn is that firms are not locating to rural areas, for example, because the cost savings in the form of cheaper labor and land rent cannot offset the benefits of urban locations.

This ignores the importance of external dis-economies.[22] While there has been empirical success in measuring agglomeration effects through modelling, Carlino (1979) and others correctly point out that these efforts at best provide a gauge of **net** benefits that accrue to firms selecting a particular metropolitan

[22]Nourse defines external dis-economies as "upward shifts of the average-cost curves of each firm as the whole industry expands in one place" (p. 89). External dis-economies result for a variety of reasons including increase in raw material cost caused by longer transport distances, higher production costs because the increased demand for inputs exerted by firms drives up input price, increased land rent caused by population.

site. The clear delineation of the external dis-economies and economies of firm location decisions is needed before these decisions can be analyzed as optimal.

The singular choices that maximize the individual welfare of firms is not necessarily socially desirable. This last point is critical. The private costs and benefits of firm location are not necessarily those borne and received by society. Beyond measuring private agglomerative economies is the task of discerning the social economies and dis-economies of firm location. Extending the familiar example of a pollution externality to the problem of firm location, a firm may locate in a city and produce specialized inputs for an industry which results in lower average costs of production for all firms using the input. But this firm may emit pollution at a level that exceeds the costs savings it brings to the industry. Although the industry realizes these costs savings, it does not compensate society for the pollution that accompanies its location decision.

External costs may be direct or indirect. Westaway (1974) discusses several costs frequently ignored in the social accounting of location decisions. The spatial concentration of firm headquarters in cities can have a sizable impact on the access that non-metropolitan laborers have to management, provoking a loss of community control over business and inefficiencies in production. Consigning less-skilled laborers to the rural areas or the fringes of the city center while white-collar labor concentrates in suburban areas implies that differing levels of access to education, transportation, and health care emerge. As Westaway comments, "(T)he most relevant question here is the way in which regional variations in the occupational distribution of labor affect social mobility" (p. 152).

An accurate measurement of external economies and dis-economies allows the consideration of whether there exists an optimal city size[23] which is less than the concentration seen in most urban areas. These efforts have been modest. Edel (1972) attempted to measure the dis-economies of congestion through decreases in land values. The author estimated regressions of city land values as a function of population and other explanatory variables including a city's tendency to host corporate headquarters. By separating corporate cities from non-corporate, the author developed a measure of land values for two city types. His findings indicate that on average, agglomeration economies surpass dis-economies for cities up to one-half of a million in population but that for larger cities, "the balance is more doubtful." (p. 62). For non-corporate cities over a million in population, the gains from agglomeration did not tend to overshadow the costs of urban congestion.

Edel also found an unequal distribution of the losses due to congestion amongst agents in society. He concludes that the distribution of capital gains and losses due to city growth is regressive for several reasons. First, the development of cities causes a fragmentation between the inner urban areas and suburbs. Much of the cost of metropolitan services falls to the poorer worker who must locate close to work because of high commuting costs. Second, Edel maintains that regressive distribution develops because the transportation systems shuttling people from the outer regions of the metropolis to the city for work impose negative externalities on inner city residents. Also, congestion may cause deterioration of property and abandonment in the inner ring, lowering the property values of the poor. For more skilled workers, their specialized labor may not be as mobile as neoclassical theory implies and laborers must bear the congestion costs even if the wage compensation is not adequate. These costs may come in the form of sizable commuting costs and lost time.

---

[23]Optimal size is defined in the neoclassical tradition to be where marginal social costs are equal to marginal social benefits.

VIII. Summary and Conclusions

This paper explores the role of agglomeration economies in describing the locational decisions of firms as a means to understand the recent migration from rural areas to urban and the related rural development policy implications. Agglomeration is considered with respect to other factors. For example, search costs and non-economic considerations may play a key role in determining location. Also, the evaluation of industries and firms within the changing context of an economy shifting from goods-producing to services-producing, national to global, and labor-intensive to technology-intensive manufacturing, may also contribute to the observed location choices. The empirical work explicitly reviewed here has concentrated on economies of scale that are external to the firm and internal to the industry (localization economies) and economies of scale that are external to both the firm and the industry (urbanization economies).

The first step in measuring agglomeration economies is to confirm their existence. Methods include surveys, indices, factor analysis, input-output technical coefficients, as well as regression analysis. The general approach was to link spatial association with economic linkages and agglomeration. While results varied in accordance with the type of agglomeration measure used, the general conclusion is that agglomeration economies play a part in explaining firm and industry clusters.

Many studies use production function models to measure the extent of agglomeration economies. This work generally begins with either the Cobb-Douglas or the CES production function. Two ways of measuring agglomeration economies are the Hicks-neutral shift and returns-to-scale parameters. Whether time-series or cross-sectional analysis is more appropriate is a point of debate. The former controls for city variation while the latter controls for technological change. To circumvent the problems associated with the measurement of capital some authors have employed the Dhrymes formulation. The overall results are encouraging in that both localization and urbanization economies are found to exist in varying degrees. Population appears to be a viable proxy for urbanization economies while localization economies vary across industrial sectors.

Urban land use theory employs agglomeration economies to explain spatial land patterns. Researchers in this area have assumed the existence of "spatial externalities" analogous to the concept of agglomeration economies used by economists. This theory directly relates productivity of a firm to the proximity of other firms and households via communication and transaction costs. While there has been little empirical testing of this theory, its significance lies with the applicability of agglomeration economies as a basic tenet.

Finally, dis-economies of agglomeration are considered to estimate an optimal city size. These costs have been difficult to isolate. It has been argued that dis-economies are actually included in the measurements of localization and urbanization economies; i.e., the measures of these factors are net of any dis-economies. Work that has considered these costs directly has focused on land values and congestion. While these costs may be substantial, conclusions based on the observed location patterns suggest that the positive effects of agglomeration outweigh the negative.

Whether estimating an optimal city size or directly measuring agglomeration economies, the distinction between urbanization and localization economies is often overlooked. These two categories of agglomeration economies are strikingly dissimilar and their relative strengths dictate different policy directions. If the bulk of agglomeration benefits a firm experiences come from the intrinsic savings of a city location, efforts to lure industry to the rural areas will likely require a substantial level of subsidy to compensate for this loss and may be unsuccessful. But if firms receive benefits not so much from a metro site but because these areas are home to other linked industries, then if interdependent firms were transported *en masse* from a city to a rural location, their cost structures would not be adversely affected. The relative composition of agglomeration benefits is an important policy question that has been ignored.

Revitalizing the rural economy is increasingly an important component of the political and social agendas. Concomitant to this is a public desire to ease the high costs of urbanization. But to date, efforts to move industry out of densely populated areas via economic development programs have enjoyed only modest success (Schmid, 1991). Returning to the fundamental reasons why firms locate in and around metropolitan areas may offer suggestions about the way in which a more socially optimal population distribution can be attained.

Agglomeration research has promise in answering these questions but has tended to go in directions that are analytically interesting but not very useful. Redirecting efforts in this area should yield work that is important in helping regional scientists define a research agenda that is shifting with changes in rural life. Much of the future needs of rural areas will not be agricultural or agrarian based. Recognizing that there is a need to assist in this movement in emphasis from farming to other kinds of enterprises is part of the new responsibilities of economic researchers interested in rural areas.

Appendix: Derivation of Dhrymes Specification

The generalized form of the CES production function is:

$$Q=A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\frac{\theta}{\rho}} \tag{18}$$

where:

Q	=	Output
L	=	Labor
A	=	Efficiency parameter
δ	=	Share of income going to each factor ($0 \leq \delta \leq 1$)
ρ	=	Substitution parameter
θ	=	Degree of homogeneity parameter

Dhrymes shows that if perfect competition is assumed in the input market, the CES model collapses to:

$$\ln(L)= \frac{-\rho}{\theta(1+\rho)} \ln(A) - \frac{1}{(1+\rho)} \ln(W) + \frac{\theta+\rho}{\theta(1+\rho)} \ln(Q) \tag{19}$$

The derivation is as follows:

$$Q = \frac{A}{[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{\frac{\theta}{\rho}}} \tag{20}$$

$$\frac{Q}{L} = \frac{A}{[L^{\frac{\rho}{\theta}}(\delta K^{-\rho} + (1-\delta)L^{-\rho})]^{\frac{\theta}{\rho}}} \tag{21}$$

$$\frac{Q}{L} = \frac{A}{[L^{\frac{\rho}{\theta}-\rho}(\delta(\frac{K}{L})^{-\rho} + (1-\delta))]^{\frac{\theta}{\rho}}} \tag{22}$$

$$\frac{Q}{L} = \frac{AL^{\theta-1}}{[\delta(\frac{K}{L})^{-\rho} + (1-\delta)]^{\frac{\theta}{\rho}}} \tag{23}$$

Assuming perfect competition, (18) is differentiated with respect to labor:

$$W = \frac{\theta A\rho(1-\delta)}{\rho[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{\frac{\theta}{\rho}+1} L^{\rho+1}} \tag{24}$$

Simplifying:

$$W = \frac{\theta(1-\delta)A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\frac{\theta}{\rho}}}{[\delta K^{-\rho} + (1-\delta)L^{-\rho}]L^{\rho+1}} \tag{25}$$

$$W = \frac{\theta(1-\delta)Q}{[\delta K^{-\rho} + (1-\delta)L^{-\rho}]L^{\rho+1}} \tag{26}$$

$$W = \frac{Q\ \theta(1-\delta)}{L[\delta K^{-\rho} + (1-\delta)L^{-\rho}]L^{\rho}} \tag{27}$$

$$W = \frac{Q\ \theta(1-\delta)}{L[\delta(\frac{K}{L})^{-\rho} + (1-\delta)]} \tag{28}$$

$$[\delta(\frac{K}{L})^{-\rho} + (1-\delta)] = \theta(1-\delta)(\frac{Q}{L})(\frac{1}{W}) \tag{29}$$

Substituting (27) into (21):

$$(\frac{Q}{L}) = AL^{\theta-1}[(\frac{Q}{L})(\frac{1}{W})\theta(1-\delta)]^{-\frac{\theta}{\rho}} \tag{30}$$

$$(\frac{Q}{L}) = AL^{\theta-1}L^{\frac{\theta}{\rho}}Q^{-\frac{\theta}{\rho}}W^{\frac{\theta}{\rho}}[\theta(1-\delta)]^{-\frac{\theta}{\rho}} \tag{31}$$

$$Q = AL^{\theta+\frac{\theta}{\rho}}Q^{-\frac{\theta}{\rho}}W^{\frac{\theta}{\rho}}[\theta(1-\delta)]^{-\frac{\theta}{\rho}} \tag{32}$$

$$L^{-(\theta+\frac{\theta}{\rho})} = AQ^{-(\frac{\theta}{\rho}+1)}W^{\frac{\theta}{\rho}}[\theta(1-\delta)]^{-\frac{\theta}{\rho}} \tag{33}$$

$$-(\frac{\theta(\rho+1)}{\rho})\ln(L) = \ln(A)-(\frac{\theta+\rho}{\rho})\ln(Q)+(\frac{\theta}{\rho})\ln(W)-(\frac{\theta}{\rho})\ln[\theta(1-\delta)] \tag{34}$$

$$\ln(L) = -\frac{\rho}{\theta(1+\rho)}\ln(A)+\frac{\theta+\rho}{\theta(1+\rho)}\ln(Q)-(\frac{1}{1+\rho})\ln(W)+(\frac{1}{1+\rho})\ln[\theta(1-\delta)] \tag{35}$$

Bibliography

Barkley, David L., and Kevin T. McNamara, "Manufacturers' Location Decisions: Do Surveys Provide Helpful Insights?", *International Regional Science Review*, forthcoming.

Begovic, Boris, "Industrial Diversification and City Size: The Case of Yugoslavia," *Urban Studies*, 29(1), February 1992, pp. 77-88.

Bergsman, Joel Peter Greenston and Robert Healy, "The Agglomeration Process in Urban Growth," *Urban Studies*, 9, (1972), pp. 263-288.

Blair, John P., and Robert Premus, "Major Factors in Industrial Location: A Review," *Economic Development Quarterly*, 1(1), 1987, pp. 72-85.

Bopp, Richard and Peter Gordon, "Agglomeration Economies and Industrial Economic Linkages: Comment," *Journal of Regional Science*, 17(1), 1977, pp. ??.

Calem, Paul S. and Gerald A. Carlino, "Urban Agglomeration Economies in the Presence of Technical Change," *Journal of Urban Economics*, 29(1), January 1991, pp. 82-95.

Carlino, Gerald R., "Manufacturing Agglomeration Economies as Returns to Scale: A Production Function Approach" *Papers of the Regional Science Association*, 50, 1982, pp. 95-108.

Carlino, Gerald A. and Edwin S. Mills, "The Determinants of County Growth," *Journal of Regional Science*, 27(1), 1987, pp. 39-53.

Carlton, Dennis W., "Why Do New Firms Locate Where They Do: An Econometric Model," in *Interregional Movements and Regional Growth*, Edited by William C. Wheaton, The Urban Institute, 1979.

Dorf, Ronald J. and M. Jarvin Emerson, "Determinants of Manufacturing Plant Location for Nonmetropolitan Communities in the West North Central Region of the U.S.," *Journal of Regional Science*, 18(1), 1978, pp. 109-120.

Dhrymes, P.J., "Some Extensions and Tests for the CES Class of Production Functions," *Review of Economics and Statistics*, 47: 357-366, 1965.

Duffy, Neal E., "Returns to Scale Behavior and Manufacturing Agglomeration Economies in U.S. Urban Areas," *Regional Science Perspectives*, 17(1):42-54, 1987.

Edel, Matthew "Land Values and the Costs of Urban Congestion: Measurement and Distribution," in *Political Economy of Environment: Problems of Method*, Papers presented at the symposium, Maisons des Sciences de l'Homme, Paris, École Pratique des Hautes Études and Mouton and Co., France, pp. 61-90, 1972.

Fujita, Masahisa, "Existence and Uniqueness of Equilibrium and Optimal Land Use," *Regional Science and Urban Economics*, vol. 15, 1985, pp. 295-324.

Fujita, Masahisa, "Towards General Equilibrium Models of Urban Land Use," *Revue Economique*, vol. 1, 1985, pp. 135-167.

Greytak, David and Paul Blackley, "Labor Productivity and Local Industry Size: Further Issues in Assessing Agglomeration Economies," *Southern Economics Journal*, 51(4): 1121-1129, 1985.

Grimaud, A., "Agglomeration Economies and Building Height," *Journal of Urban Economics*, vol. 25, 1989, pp. 17-31.

Hamilton, Joel R., Norman K. Wittlesey, M. Henry Robison, and John Ellis, "Economic Impacts, Value Added, and Benefits in Regional Project Analysis," *American Journal of Agricultural Economics*, 73(2): 334-344, May 1991.

Hansen, Eric R. "Agglomeration Economies and Industrial Decentralization: The Wage-Productivity Trade-offs," *Journal of Urban Economics*, vol. 28, 1990, pp. 140-159.

Hartshorn, Thomas A. *Interpreting the City: An Urban Geography*, John Wiley and Sons, 2nd edition, 1992.

Helsley, Robert W., and William C. Strange, "Agglomeration Economies and Urban Capital Markets," *Journal of Urban Economics*; 29(1): 96-112, 1991.

Hewings, Geoffrey J.D., and Rodney C. Jensen, "Regional, Interregional and Multi-regional Input-Output Analysis," in *Handbook of Regional and Urban Economics*, Peter Nijkamp, ed., Volume 1, Elsevier Science Publishers B.V, 1987, pp. 295-353.

Hoover, Edgar and F. Giarratani, *An Introduction to Regional Economics*, Third Edition, Alfred A. Knopf, pubs, 1984 pp. 119-125.

Isard, Walter, *Location and Space Economy*, Cambridge, Mass: M.I.T. Press, 1956. pp. 182-188.

Kawashima, Tatsuhiko, "Urban Agglomeration Economies in Manufacturing Industries," *Papers of the Regional Science Association*, 7: 80-91, 1977.

Knapp, Thomas A. and Philip A. Graves, "On the Role of Amenities in Models of Migration and Regional Development," *Journal of Regional Science*, vol. 29, no.1, 1989, pp. 71-87.

Krugman, Paul, *Geography and Trade*, Leuven University Press, Leuven, Belgium, 1991.

Lever, W.F., "Industrial Movement, Spatial Association and Functional Linkages," *Regional Studies*, 6: 371-384, 1972.

Luke, Jeffrey S., Curtis Ventriss, B.J. Reed, and Christine M. Reed, *Managing Economic Development: A Guide to State an Local Leadership Strategies*, Jossey-Bass Publishers, 1988.

Maki, Wilbur, *Reshaping the City Region in Global Competition*, Dept. of Agricultural and Applied Economics Staff Paper P92-26, University of Minnesota, St. Paul, Minnesota, 1992.

Marcus, Matityahu, "Agglomeration Economies: A Suggested Approach," *Land Economics*, 41(3): 279-284, 1965.

Meyer, David R. "Agglomeration Economies and Urban Industrial Growth, A Clarification and Review of Concepts," *Regional Science Perspectives*, 7(1): 80-92. 1977.

Minneapolis Star and Tribune "Economic Realities," Peter Leyden, p. 1, 6/28/93.

Minnesota Average Covered Employment and Wages, Department of Jobs and Training, 1991, p. 3 and 29.

Moomaw, Ronald L., "Productivity and City Size: A Critique of the Evidence," *Quarterly Journal of Economics*, 94: 675-688, 1981.

Moomaw, Ronald L., "Is Population Scale a Worthless Surrogate for Business Agglomeration Economies?" *Regional Science and Urban Economics*, 13: 525-545, 1983.

Moomaw, Ronald L., "Firm Location and City Size: Reduced Productivity Advantages as a Factor in the Decline of Manufacturing in Urban Areas", *Journal of Urban Economics*, 17: 73-89, 1985.

Moomaw, Ronald L., "Have Changes in Localization Economies Been Responsible for Declining Productivity Advantages in Large Cities?", *Journal of Regional Science*, 26(1): 19-32, 1986.

Moses, Leon N., "Location and Theory of Growth," *Quarterly Journal of Economics*, 72: 259-272, 1958.

Nakamura, Ryohei, "Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities," *Journal of Urban Economics*, 17: 108-124, 1985.

Nourse, Hugh O., *Regional Economics: A Study in the Economic Structure, Stability and Growth of Regions*, McGraw-Hill Book Company, New York, 1968.

Ogawa, Hideaki and Masahisa Fujita, "Equilibrium Land Use Patterns in a Nonmonocentric City," *Journal of Regional Science*, vol. 20, no. 4, 1980, pp. 455-475.

Papageorgiou, Y. Y. and J. F. Thisse, "Agglomeration as Spatial Interdependence between Firms and Households," *Journal of Economic Theory*, vol. 37, 1985, pp. 19-31.

Pascal, Anthony H. and John. J. McCall, "Agglomeration Economies, Search Costs, and Industrial Location," *Journal of Urban Economics*, 8(3): 383-88, 1980.

Peterson, Willis, "Production and Supply: 101 Questions and Answers," mimeo, Department of Agricultural and Applied Economics, University of Minnesota, St. Paul, 1990.

Schmenner, Roger W., "Look Beyond the Obvious in Plant Location," *Harvard Business Review*, 52: 126-132, 1979.

Schmid, Allan A., "Rethinking Cost in Population Settlement Patterns," *Choices*, Second Quarter, 1991, p. 39.

Scott, Allen J., "The Spatial Organization of a Local Labor Market: Employment and Residential Patterns in a Cohort of Engineering and Scientific Workers," *Growth and Change*, 23(1): 94-115, 1992.

Segal, David "Are There Returns to Scale in City Size?," *Review of Economics and Statistics*, 58, 339-350, 1976.

Shefer, Daniel, "Localization Economies in SMSA's: A Production Function Analysis," *Journal of Regional Science*, 13(1): 55-64, 1973.

Shukla, Vibhooti and Oded Stark, "Policy Comparisons with an Agglomeration Effects-Augmented Dual Economy Model," *Journal of Urban Economics*; 27(1): 1-15, 1990.

Stanback, Thomas Jr., *The New Suburbanization: Challenge to the Central City*, Westview Press, Boulder, CO 1991, p. 79.

Statistical Abstract of the United States, U.S. Department of Commerce, Bureau of Census, 110th edition, 1990.

Streit, M.E., "Spatial Associations and Economic Linkages Between Industries," *Journal of Regional Science*, 9(2): 177-188, 1969.

Sveikauskas, Leo, "The Productivity of Cities," *The Quarterly Journal of Economics*, 89(3): 393-413, 1975.

Tabuchi, Takatoshi, "Urban Agglomeration, Capital Augmenting Technology, and Labor Market Equilibrium," *Journal of Urban Economics*, vol. 20, 1986, pp. 211-228.

Taylor, M.J., "Local Linkage, External Economies and the Ironfoundry Industry of the West Midlands and East Lancashire Conurbations," *Regional Studies*, 7: 387-400, 1973.

Thorngren, B., "How Do Contact Systems Affect Regional Development?," *Environment and Planning*, 2: 409-427, 1970.

United Nations Chronicle, "Half the World Will Live in Cities by the Year 2000," 23: 92, August 1986.

Weber, Adna F., *The Growth of Cities in the Nineteenth Century*, Columbia University Studies in History, Economies, and Public Law, 11 New York Macmillan, 1899. rev. ed, Ithaca, N.Y. Cornell University Press, 1963.

Weber, Alfred *Theory of the Location of Industry*, translated by C. J. Friedrich, Chicago, Illinois, The University of Chicago Press, 1956.

Westaway, J., "The Spatial Hierarchy of Business Organizations and its Implications for the British Urban System," *Regional Studies*, 8: 145-155, 1974.

Wheaton, William C. and Hisanobu Shishido, "Urban Concentration, Agglomeration Economies, and the Level of Economic Development," *Economic Development and Cultural Change*, 30(1): 17-30, 1981.

Zipf, George Kingsley, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, Mass., 1949.