Spatial Distribution of Economic Activities in North America

Thomas J. Holmes

University of Minnesota and Federal Reserve Bank of Minneapolis and National Bureau of Economic Research

holmes@econ.umn.edu

John J. Stevens

Board of Governors of the Federal Reserve System

John.J.Stevens@frb.gov

June 27, 2003

Abstract

In this chapter we discuss the data sources and methods available for studying the spatial distribution of economic activity in North America. We document facts about the specialization of states and regions, as well as locations differentiated by their degree of urbanization. We also report characteristics of the industries in which locations specialize. For example, establishment size and materials intensity are shown to vary in systematic ways with regional specialization. With these facts as a backdrop, we begin to consider the question, Why do locations specialize as they do?

Keywords: geographic concentration; spatial distribution; regional specialization; establishment size

JEL Classification: L11, R12, R30

1 Introduction

How is economic activity distributed across space? This chapter addresses this question by examining data for the North American countries of the United States and Canada.[1] The chapter complements the companion studies on Europe and Asia by Combes and Overman and Fujita et al. in this volume.

The titles and geographic specificity of these commissioned chapters harken back, at least as far as economists are concerned, to an earlier era of research. We could certainly imagine the economist Edgar M. Hoover writing a treatise with the title of this chapter in the 1950s. We expect that this work, like Hoover (1948), would be full of maps and tables that attempt to explain where activity is located and why. The first two volumes of the *Handbook of Regional and Urban Economics*, published in 1986, contain nothing like this. Neither volume has a map, and neither volume has a table with raw data on the location of economic activity. In contrast, this chapter and others in this volume are loaded with maps and tables describing the distribution of economic activity.

The resurgence of interest by economists in the facts about how economic activity is distributed across space is a welcome development given the magnitude of the questions at stake. These questions include: How important are geographic considerations for determining the economic success of a location? What are the economic consequences of the pervasive use of government policies to influence the location of economic activity, such as the large subsidies state and local governments have paid in recent years to attract new automobile plants? Can certain regions potentially lose from interregional trade as core industries like manufacturing abandon some regions and concentrate elsewhere?

Some progress has been made on the theoretical front in addressing these types of questions. The 1990s witnessed a burst of theoretical activity on economic geography. According to Fujita et al. (1999), this body of work is beginning to enter a mature stage in which it is time to synthesize and summarize the work in monographs. (See also Fujita and Thisse (2002) and the chapter by Duranton and Puga in this volume). Complementing this theoretical literature is an emerging empirical front. The empirical literature (discussed in this

[1] Unfortunately, data considerations precluded us from including Mexico in the analysis.

volume in the chapter by Rosenthal and Strange) has progressed toward, but not reached, its mature phase. The quantity of empirical effort has been small in comparison with the theoretical work. In this circumstance, we think it is useful to have handbook chapters, like this one on North America and the others on Europe and Asia, that describe data sets and lay out basic facts. We expect that there will be much work in this fertile and important area in the years to come, and we hope that these chapters will be a useful resource for such work.

The focus of this essay is on the *composition* of economic activity across space. (We leave issues about the overall level of economic activity, e.g., the formation of cities, to other chapters; in particular, see the chapter by Gabaix and Ioannides in this volume.) We define a location as *specializing* in a particular industry if that location's share of employment in the industry exceeds the national average. For example, Michigan specializes in automobile production because the share of Michigan's employment that is in automobile manufacturing is approximately seven times the national share. We focus on employment to measure specialization because excellent, publicly available data exist with information on employment at the establishment level. Information about other measures of activity, such as sales or value added, is not as readily available. We provide an in-depth discussion of the potential data sources in Section 2.

We document in Section 3 that, for industries producing nontradable goods or services like retail, there is little specialization, while for tradable goods like manufactures, mining output, and agricultural products, there is a substantial amount of specialization across regions. As late as the 1950s, manufacturing activity in the United States was heavily concentrated in the Northeast and Upper Midwest around the Great Lakes, in an area often called the *manufacturing belt*. In the ensuing decades, U.S. manufacturing moved out of this northern region and into other parts of the country. We show that as of 2000, the area once delineated as the manufacturing belt in fact no longer specializes in manufacturing. In contrast, certain areas of the South have become quite specialized in manufacturing, in effect fashioning a new manufacturing belt.

We also document in Section 3 the extent of specialization among narrowly defined industries. A problem with the narrow definition of industries is that the definition may encompass only a few establishments, even at the national level. (In some cases we have as few as 10!) In such cases, the "dartboard" effect of randomness emphasized by Ellison and Glaeser (1997) potentially plays a role; we therefore use their methodology to try to control for the dartboard effect.

The new economic geography literature places great emphasis on the relation between specialization and market size. In Section 4 we report that specialization varies to a substantial degree with a particular measure of market size—the concentration of population. Obviously, specialization in farming varies inversely with the concentration of population. But, less obviously, as of 2000, rural areas specialize in manufacturing, in contrast to earlier times when manufacturing was an urban activity. Even when we exclude rural areas and look only across urban areas, we find substantial variation in specialization based on market size.

An important goal of economic research in this area should be to determine the source of specialization patterns. We have made our job easy from the start by not setting this as the goal for the chapter, but Sections 5 and 6 do move in this direction. Section 5 discusses facts about variables that are consistently found to be correlated with industry concentration. For example, one robust fact is that industries with larger plant sizes tend to have higher levels of geographic concentration. Another robust fact is that within an industry, areas of industrial concentration tend to have larger plants. Theories that try to explain concentration should be consistent with these facts. Section 6 briefly discusses several forces that are thought to determine the location of economic activity. We illustrate these forces using two familiar examples from the economic geography literature—the beverage industry (a classic "weight-gaining" industry) and the joint-location of sugar beet processing facilities and sugar beet farms (a classic "weight-losing" industry). Interpreting these classic examples through the lens of the new economic geography literature (i.e., scale economies) reveals new insights into these industries.

2 Data

We begin with a discussion of U.S. data and then turn to a brief discussion of Canadian data. The ability to conduct cross-country analysis with North American countries has recently been enhanced by the joint adoption of the North American Industry Classification System (NAICS) by the United States, Canada, and Mexico. Before the adoption of NAICS, cross-country analysis was difficult because each country had its own classification system. In addition to having greater harmonization across countries, NAICS has greater classification detail about new and emerging industries, such as services, compared with the Standard Industrial Classification (SIC) system that it replaced in the United States. Further details about the data are available in the Appendix.

2.1 United States

This subsection begins with a discussion of the County Business Patterns (CBP) data published by the U.S. Census Bureau. These data are the primary information that we use to summarize location patterns in the United States, but they do not include information about government employment, sole proprietorships, or agriculture (crops and farm animals), so we also discuss sources for this information. We conclude by mentioning other sources of data on location patterns.

The CBP data set contains information about the location of the approximately 7 million establishments with employment in the United States. The source of these data is the Census Bureau's "Business Register" (formally called the *Standard Statistical Establishment List*). The Census Bureau updates the Business Register by conducting an annual survey of multi-establishment firms, called the *Company Organization Survey*, and by incorporating various administrative data that it receives from other government agencies, including the Internal Revenue Service.

The raw, establishment-level micro data are released annually in the form of cell counts by employment-size class, industry, and location. The employment-size classes are relatively narrowly defined (Table 1). Large establishments, defined as those with 1,000 or more employees, are broken up into four relatively narrow categories. Other Census Bureau programs have only one category for establishments in the 1,000-plus category. The industry information is the six-digit NAICS code, which is the highest degree of detail for establishment classification. The location information is the county identifier.[2]

Figure 1 illustrates the kind of information that is available with the CBP cell counts. It plots the location of the 19 brewery establishments (NAICS industry 31212) in the United States with 500 or more employees in the 1999 CBP. The mapping program randomly assigns an establishment in a given county to some point within the polygon defined by the county boundaries. The resulting figure is called a dot density graph. County boundaries encompass a much larger area in the western states.

The small dots denote plants with 500–999 employees; the large dots represent the three plants with more than 1,000 employees. As mentioned, the CBP data break up the 1,000-plus size class into more detailed groupings. Only one plant is in the 2,500–4,999 employment-size range, and it is located in Jefferson County, Colorado. Publicly available data sources can be used to establish that this is the Coors brewery in Golden.[3] Only one plant is in the 1,500–2,499 category, the Anheuser-Busch plant in St. Louis, Missouri. It is against the law for the Census Bureau to present data that would disclose the operations of an individual employer. However, the Bureau does not consider cell count information to be a disclosure and has published cell counts by employment-size class for many decades. According to Census Bureau documentation (U.S. Department of Commerce, Bureau of the Census (2002a), p. vii), "The number of establishments in an industry classification and the distribution of these establishments by employment-size class are not considered to be disclosures, and so this information may be released even though other information is withheld."

[2] There is also ZIP code information, but we focus on county data. Although ZIP codes present a finer level of geographic detail than counties (there are 40,000 ZIP codes compared to 3,000 counties), they are not as clean a geographic unit as counties. Some locations have multiple ZIP codes, and the definitions change over time according to the U.S. Postal Service's administrative needs. Another issue with the ZIP code data is that the establishments with more than 1,000 employees are aggregated into a single employment-size class rather than the four employment-size classes used for the county data.

[3] For example, the U.S. Environmental Protection Agency's Toxic Release Inventory (TRI) data provide address and industry information for manufacturing plants with toxic releases. (See the Appendix.) The Coors plant is the only brewery in this data set in Jefferson County.

Because of nondisclosure requirements, the Census Bureau routinely withholds data from the information it publishes regarding total employment and payroll aggregated across establishments in the same county or state and industry. The Census Bureau requires at least three establishments in an industry/county pair before it will report total employment, but it withholds data even for industry/county pairs with hundreds of establishments. Overall, approximately 20 percent of all establishments are in industry/county pairs with withheld data. The nondisclosure issue is a severe problem if one needs to know the exact employment at a particular location. For example, if a city planner were interested in the growth in brewery employment in St. Louis from 1998 to 1999, he or she would find two establishments in this industry/county pair and would find that total employment is not disclosed. Various means to try to extract the nondisclosed employment data have been proposed, including techniques that exploit the information in the published state-level totals and further aggregated industry totals. (See Gardocki and Baj (1985).) Despite these efforts, the CBP would be of limited value in determining the exact employment growth in St. Louis breweries from 1998 to 1999.

Our interest is in documenting and understanding the location patterns of industries, and for this purpose the complete disclosure of the raw micro data, in terms of cell counts, is of substantial value. To begin with, some exercises involve combining plants into size groupings anyway, so there is literally no loss from using the public-release micro data rather than the actual Business Register. (Figure 1 is an example; see also Dinlersoz (forthcoming).) For other exercises one might need an estimate of industry employment at a particular location. With the public-release micro data, one can straightforwardly calculate an estimate of total employment across establishments by using the mean employment in the establishment's size class from Table 1 as an estimate of establishment employment and then aggregating (Holmes and Stevens (2002), Holmes (1998)).[4] We use this procedure here; further details are available in the Appendix.

To illustrate the usefulness of these estimates for our purposes, consider the Gini coefficient for the 33 three-digit NAICS industries for which employment information is completely

[4]Some researchers use the midpoint of the cell's range. However, the size distribution is skewed, so the mean is preferred. The mean has to be estimated for size classes above 1,000 (see the Appendix).

disclosed in every state (there are 50 remaining three-digit NAICS industries that have some data withheld). The Gini coefficient (discussed below) is calculated at the state level in the first column of Table 2. The second column constructs the Gini using the publicly available establishment-level size class data to estimate establishment employment. The Gini calculated with the estimates is virtually the same as with the actual data, even though for some states and industries the error in the employment estimate is large, as high as 20 percent in some cases. The comparison in Table 2 shows that the establishment-level data provide an excellent approximation to the results we would obtain if we were to use the confidential Business Register data at a secure Census Bureau facility. We also note that a useful property of the CBP data is that they are available in computer readable files from 1974 to the present.[5]

Although the industry coverage of the CBP is broad, it does not include information about farm establishments, government establishments, or sole proprietorships (self-employment). If one is interested in these forms of economic activity, then other data sources are required. The U.S. Department of Commerce's Bureau of Economic Analysis publishes this information under the Regional Economic Information System at the county and detailed sector levels. Data-withholding issues exist in some cases, but we already have data on private-sector employers from CBP, so all we need is government employment, agriculture employment, and proprietorships. For these three sectors, disclosure is not an issue.

The CBP does not provide data on output measures such as sales or value added. This information is collected by the Economic Census, which is conducted every five years. Henderson (1988) is an example of a study on productivity that uses the published Economic Census data. A difficulty with these data is that information for certain location/industry combinations is withheld. For sectors other than manufacturing, a complete release of cell counts is available by sales-size categories for ZIP codes. However, the size categories and industries are highly aggregated, so the cell counts are not as useful as the CBP release.

An important development in recent years is the increased access by outside researchers

[5]Before 1974 the CBP data are available on paper; they are used by Glaeser et al. (1992). A change was made in 1974 regarding reporting units. The methodology since then has been the same (except for the switch from SIC to NAICS in 1998).

to the confidential micro data via the Census Bureau's Center for Economic Studies (CES) and the related Resource Data Centers (RDCs). Access to these data means it is now possible to study questions that are impossible to answer with the public-establishment-level CBP data. One can examine establishment-level productivity and the link with agglomeration (Henderson (2003)). One can link different establishments over time and examine establishment growth (Dumais et al. (2002)). Or one can link establishments of the same firm (Holmes (2002)). The ability to conduct such studies makes the increased access to data through the CES and RDCs an important development.

The Census Bureau data have limitations. Although one can do some things only in the Census, the reverse is also true: Some things can be done outside the Census that cannot be done inside. For example, the Census Bureau is careful about disclosure. One cannot use the Business Register to make a map like that in Figure 1. Nor can one write a paper with access to the Business Register and mention anything about Coors or Anheuser-Busch. If one is interested in exploiting establishment links across time or within the firm, or if one wishes to study plant-level productivity, one must obtain access to the confidential data. If instead one is interested in studying the distribution of employment across counties, then given the disclosure issues, not to mention the substantial expense in time and money for the Census Bureau, the use of public CBP data is probably the preferred option.

The CBP and Economic Census data discussed so far are data collected at the *establishment* level. The Census Bureau also publishes detailed geographic data based on data collected at the *individual* level as part of the decennial census. We focus on data collected at the establishment level because industry definition is generally more detailed and more reliable than it is for data collected at the individual level. Furthermore, in part of our analysis we exploit the information on establishment size in the CBP data, information that is not available in the population data. We use the establishment-level data here, but the population data are also a rich source for information on the geographic distribution of economic activity.

In addition to the Census Bureau, private companies collect information about establishment location. Dun & Bradstreet is perhaps the best known of these firms. Its data have been used recently by Rosenthal and Strange (2001) and earlier by Evans (1987) and others. An advantage of the Dun & Bradstreet data is that they are more accessible than the confidential Census Bureau data. However, "The data are not collected or organized with scientific research in mind, and there are consequently a variety of potential problems" (MacDonald (1985), p. 173). In particular, small firms tend to be underrepresented. Moreover, there are problems with industry definitions.[6] If one is conducting a cross-section analysis of the location of employment across counties or ZIP codes, then the public, establishment-level CBP data are superior to the Dun & Bradstreet data because the CBP data are free of charge and are of higher quality. If, instead, one needs to link establishments across time (e.g., to look at firm births), and if access to the confidential micro data is unavailable, then the Dun & Bradstreet data are a useful alternative (Rosenthal and Strange (2001)).

2.2 Canada

StatCan, the Canadian statistical agency, conducts the Canadian census every five years for both population and business. It creates a database similar to the U.S. CBP database known as the Canadian Business Patterns. The employment-size cell counts are not as fine; 500 and above is a single category. We use the data in a manner similar to that for the U.S. CBP data.

StatCan data are costly relative to the U.S. CBP data. The U.S. Census Bureau places its data in the public domain, where it can be reproduced free of charge. A compact disk (CD) containing the data sells for $50. In contrast, the Canadian data are protected as intellectual property. The Canadian CD costs $1,000 with an academic discount.

3 Patterns of Regional Specialization

This section discusses patterns of regional specialization in three parts. The first part examines patterns for broad sectors and broad regions. Here we introduce some common measures

[6] In the Dun & Bradstreet data, administrative offices of manufacturers, for example, clearly are counted as manufacturing establishments, whereas the Census Bureau treats them by their function, i.e., a sales office is in wholesaling. This distinction may explain why the Dun & Bradstreet data consist of 670,000 manufacturing establishments, almost twice as many as recognized by the Census Bureau.

of specialization, including the employment location quotient and the locational Gini coefficient. The second part examines specialization patterns for narrowly defined industry and geographic definitions. Here the dartboard issues raised by Ellison and Glaeser (1997) are relevant, and we use their measure of concentration to classify industries. The third part gives special attention to the specialization patterns of the manufacturing sector.

3.1 Broad Sectors and Broad Regions

We begin by describing the location of economic activity for highly aggregated geographic industries and regions. The simplest measure of geographic concentration is the *location quotient,* or LQ. Suppose there are M locations indexed by i. Let s_i denote location i's share of industry employment, and let x_i denote its share of total employment. The LQ for location i is then defined as

$$LQ(i) \equiv \frac{s_i}{x_i}.$$

The LQ is therefore a ratio of a location's share of industry employment to its share of aggregate employment. Values above one mean that the location is relatively specialized in the industry, while values below one indicate that the location has less industry employment than would be predicted based on its aggregate employment share. While we typically use employment when calculating LQs, in principle we could use sales, number of establishments, or any other measure comparable across industries and regions. Tables 3 and 4 display employment LQs for major sectors of the economy by census divisions for the United States and groupings of provinces for Canada.[7] U.S. census divisions are roughly on the same order of magnitude in terms of population size—the largest is only three times bigger than the smallest. For Canada we use four groups: Ontario, Quebec, the eastern provinces, and the

[7]The nine U.S. Census divisions are New England (ME, NH, VT, MA, RI, CT), Middle Atlantic (NJ, NY, PA), East North Central (OH, IN, IL, MI, WI), West North Central (MN, IA, MO, ND, SD, NE, KS), South Atlantic (DE, MD, DC, VA, WV, NC, SC, GA, FL), East South Central (KY, TN, AL, MS), West South Central (AR, LA, OK, TX), Mountain (MT, ID, WY, CO, NM, AZ, UT, NV), and Pacific (WA, OR, CA, AK, HI).

The Canadian groupings are Quebec, Ontario, the eastern provinces (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland, and Labrador), and the western provinces (Manitoba, Saskatchewan, Alberta, British Columbia, Yukon Territory, and Northwest Territories).

western provinces including territories. Including Quebec with the eastern provinces would make for three groups with nearly identical populations, but this would mask the obviously different industrial profiles of Quebec and the eastern provinces.

The degree of geographic concentration exhibited by certain sectors in Tables 3 and 4 is striking. First consider manufacturing. The shares of employment in manufacturing in the East South Central and East North Central regions of the United States are nearly 35 percent greater than in the nation as a whole. (These regions have LQs of 1.34 and 1.36.) In contrast, the Mountain region, with a manufacturing LQ of 0.65, has a manufacturing share of employment that is 35 percent lower than the national share. The variations in agriculture and mining are even more dramatic. New England's LQ for agriculture is only 0.28 and for mining is only 0.12 (which happens to be the lowest LQ in the entire table). Manufacturing, mining, and agriculture are obviously sectors with substantial possibilities for trade across regions. Tables 3 and 4 indicate that these possibilities for trade are being realized as regions specialize.

The possibilities for trade in sectors like retail are relatively limited. Accordingly, there is little specialization across regions in this sector. The LQ ranges from a high of 1.06 to a low of 0.92. The government sector is also fairly evenly distributed across census divisions, but perhaps less so than one might expect. At the state level the government sector exhibits even more concentration; for example, the effect of federal employment in the District of Columbia and the surrounding states of Maryland and Virginia becomes apparent. The LQs for educational services in the Middle Atlantic and New England regions are high, reflecting the location of many colleges and private schools in these areas.

Canada is also characterized by a significant amount of concentration. Ontario is the business capital; it specializes in manufacturing, finance and insurance, real estate, and business service industries. Quebec is the most specialized region in manufacturing, but, excluding those industries closely tied to natural resources, it otherwise resembles the nation as a whole. The combined eastern and western provinces dominate the natural resource sectors—agriculture, fishing and trapping, logging and forestry, and mining—in terms of both relative and absolute employment.

The broad picture of concentration is unchanged if we drop the government and farm sectors (not shown); the U.S. LQs typically change by no more than 0.01. Therefore, in most of what follows, we will drop the government and farm sectors since they are not a part of the U.S. CBP data set.

LQs are useful for describing which locations specialize in an industry. But if all one wishes to know is whether geographic specialization is a characteristic of a particular industry, then a single statistic that summarizes the information contained in the industry's LQs would suffice. One such statistic is the *locational Gini coefficient*. Let i index the set of M locations ordered by decreasing values of LQ. Then this coefficient is defined as

$$Gini \equiv 1 - \sum_{i=1}^{M} x_i \left(s_i + 2 \sum_{j=i+1}^{M} s_j \right).$$

The Gini coefficient takes on values in $[0, 1]$, where a value of zero denotes equal specialization across locations and a value of one denotes extreme inequality. Kim (1995) uses Gini coefficients to show the long-run trends within U.S. manufacturing concentration. He shows that concentration increased until the turn of the twentieth century, flattened out, and then declined continuously from the 1930s until the end of his panel in 1987. Table 5 shows that at the sector level, the Gini coefficients changed very little over the last part of the twentieth century.[8] Unsurprisingly, mining and agricultural services have the highest Gini coefficients, while retail and services have the lowest.

3.2 Narrow Industries and More Detailed Geography

We now turn to more narrowly defined industries and locations. The examples of specialization of the auto industry in Detroit, of high-tech in Silicon Valley and Boston's Route 128, and of gambling in Las Vegas and Atlantic City are very familiar. Here we assess the quantitative significance of specialization in industries at this level of detail.

When we examine narrowly defined industries, naturally we will find some industries with a relatively small number of plants. In an important paper, Ellison and Glaeser (1997) argue that in industries with relatively few establishments, randomness alone can create the illusion of concentration even if the underlying process that allocates establishments to

[8] The NAICS data begin in 1998. To make the intertemporal comparison, we use SIC code data that are available for 1977 and 1997.

locations does so uniformly. They term this the dartboard problem, drawing an analogy between the clustering of establishments in space to the random clumping of darts on a dartboard. They create an index that measures the amount of clustering over and beyond that which we would expect to find based on randomness alone. In this subsection we first discuss the EG index and then apply it to the CBP data.

The EG index is defined as

$$\gamma \equiv \frac{G - \left(1 - \sum_i x_i^2\right) H}{\left(1 - \sum_i x_i^2\right)(1 - H)}, \tag{1}$$

where

$$G \equiv \sum_{i=1}^{M} (s_i - x_i)^2$$

is a measure of raw geographic concentration, s_i denotes the share of an industry's employment in each of M locations, x_i denotes the share of aggregate employment in each location, and

$$H \equiv \sum_{j=1}^{N} z_j^2$$

is the plant Herfindahl for the industry. Ellison and Glaeser show that if plants are randomly distributed across locations with probabilities given by x_i, then the expected value of this measure is zero. Thus, a positive value of γ indicates a level of concentration over and above what one would expect by chance.

If we define an industry at a highly aggregated level, such as all of manufacturing, the plant-level Herfindahl is essentially zero. In this case, the Ellison and Glaeser (1997) measure of concentration (or related measures like that proposed by Maurel and Sédillot (1999)[9])

[9] Maurel and Sédillot (1999) compute an EG-type index of the form

$$\gamma^{MS} \equiv \frac{\frac{G^{MS}}{1-\sum_i x_i^2} - H}{1 - H}$$

where

$$G^{MS} \equiv \sum_{i=1}^{M} \left(s_i^2 - x_i^2\right).$$

Thus, γ^{MS} looks just like a rescaled version of γ^{EG},

$$\left(1 - \sum_i x_i^2\right)^{-1} \gamma^{EG},$$

but with a different measure for raw geographic concentration that is discussed in their paper.

collapses down to a measure of raw geographic concentration obviating the need for the EG index. This justifies our use of the easy-to-interpret Gini coefficient in the previous subsection. In narrowly defined industries with few plants, the Herfindahl may be much greater than zero, so using the EG approach will make a difference.

We use the CBP data to calculate the EG index. The fact that this data set has establishment-level employment estimates permits us to estimate H. Ellison and Glaeser calculated their index for each manufacturing industry, using states as their benchmark geographic unit. Following their lead, we use states as our geographic unit, but we expand their analysis by calculating the index for all 1,082 six-digit NAICS industries in the CBP data.

Table 6 summarizes these calculations by major sector. The distributions are generally quite skewed as a comparison of the mean and median values of γ suggests. The median industry within manufacturing has an index of 0.020 while the mean is 0.041. With a maximum value over manufacturing industries of 0.585, it is clear that there are some highly concentrated industries. However, in about half of the major sectors, the minimum is negative. Out of the 1,082 industries, 48 have negative values for γ.

There is an open issue concerning how one should interpret a negative γ. One intuitive idea is to interpret it as indicating a less-than-random distribution of establishments. For example, suppose we throw capitol buildings at states like darts on a dartboard with a 1/50 probability of landing in each state (so that $s_i = 0.02$). The probability that we will observe exactly one capitol building in each state is extremely low. It is far more likely that at least some states will get more than one capitol building. Given the observed realization of one capitol building in each state, it is straightforward to calculate that $G = 0$ in this case and that $\gamma = -\frac{1}{49}$. In this case the negative γ gets things right at least in terms of the sign. Capitol buildings are *more* geographically dispersed than we would get from a dartboard process. This example is symmetric to the positive γ case in which a particular industry is *less* geographically dispersed than in a dartboard process.

This idea is intuitive. Some of the industries with negative values for γ are those like radio networks that we expect to be dispersed. Nonetheless, we think this issue warrants additional work, because a negative γ has no structural interpretation. In the original Ellison and Glaeser setup, the limiting case of $\gamma = 0$ is the pure dartboard model. It would be useful to generalize this structural approach to explicitly incorporate a "more dispersed than random" portion of the parameter space. In Section 6 of this chapter we return to this issue.

Table 7 displays the most and least concentrated industries for our U.S. data. Several observations are in order. In both lists, the plant Herfindahls vary greatly. A large Herfindahl is neither necessary nor sufficient for a large γ. For example, both casino hotels and scale and balance manufacturers have small Herfindahls, but only casino hotels are concentrated due to their high degree of geographic concentration. Conversely, engineered wood member manufacturers and phosphate rock mining both have relatively large Herfindahls and a large degree of geographic concentration, but only the phosphate mining is concentrated according to the EG index. Engineered wood manufacturing is an example of an industry where concentration is driven by dartboard effects. The interaction between plant concentration and geographic concentration permits a rich array of possibilities.

In addition to the negative γ issue, we mention another issue that remains open. Duranton and Overman (2002) note that the EG index, and related indices, will be downwardly biased due to the "border effect." To see how this effect arises, suppose we are computing EG indices at the state level. Then any agglomeration of economic activity that crosses the state boundary will be chopped into two pieces. For example, Gary, Indiana, is separated from Chicago, Ilinois; Kansas City, Missouri, is separated from Kansas City, Kansas; and so on. This problem increases as we move to finer geographic units such as counties. To address this issue, Duranton and Overman take a unique approach afforded them by a very detailed data set on the locations of establishments in the United Kingdom. They are able to compute, with a high degree of accuracy, the universe of pairwise distances between establishments. This allows them to look at the entire distribution of plant-to-plant distances and to measure cumulative agglomeration effects at any given distance. This same approach could be used with the CBP data or ZIP code data, which, as we have argued earlier, are a form of public release of the micro data. However, note that the plant distance measures here are only an approximation, as the county or ZIP code is known, but not the precise geographic coordinates. For studying certain issues in agglomeration, like that in the automobile parts industry where the distances involved are can be fairly large (on the order of 50 to 200 miles), plant distance measures constructed from the CBP will be satisfactory. But in other cases where we are interested in a very fine degree of geographic detail, such as where shops are located within a shopping mall, these distance measures will be unsatisfactory.

3.3 Regional Specialization in Manufacturing

In this subsection we give special focus to the location of manufacturing activity. As discussed in the first part of this section, there is substantial regional specialization in manufacturing. Here we examine how this specialization pattern has evolved over the postwar period.

We begin with a map of the location of large manufacturing plants in the United States in 1947. Each dot in Figure 2 represents the location of a single manufacturing plant with 250 or more employees. The location data are from the 1947 census and are analogous to the CBP data. (See the Appendix for a discussion of the data.) There were 10,219 manufacturing plants with 250 or more employees in 1947, and they accounted for approximately 8.5 million employees, more than half of the 14.3 million manufacturing employees.

A striking feature of this map is the high concentration of large manufacturing plants around the Northeast and Great Lakes regions. Economic geographers have referred to this region as the *manufacturing belt,* and over the years various delineations of the boundary for this belt have been proposed. In Figure 2, we follow the definition of the manufacturing belt proposed by Fritz (1943) in the *Industrial Location and Natural Resources* volume referred to in the introduction to this chapter.[10] The same map in the Fritz work identifies two other manufacturing areas that we illustrate in Figure 2. One is the Piedmont manufacturing region in the Southeast. The other is the California manufacturing zone (which is essentially a box that connects the Bay area with Los Angeles).

Note that the three demarcated areas, while having much more manufacturing activity than the rest of the country, are not each uniformly covered with dots. In the Piedmont area, the area of heaviest concentration is well within the boundary. In the manufacturing

[10] We use county boundaries to delineate the manufacturing belt border; a county is either inside the belt or outside it. In Fritz (1943), the boundary is not based on counties. The boundary we illustrate here is the best approximation we can make using counties. A more recent alternative delineation that is very similar is in Pred (1965).

belt, there is a large gap separating the Midwest from what geographers call *Megalopolis,* the area stretching from Washington, D.C., to Boston. This gap contains the Appalachian Mountains and is part of what is designated *Appalachia*.

Outside of these three regions, there are a few cities with substantial manufacturing, like Minneapolis, Houston, and Dallas. But much of the rest of the country is essentially devoid of large manufacturing plants. North Dakota, Nevada, and Wyoming have only two such plants, and New Mexico has only one plant.

Of course, states like North Dakota also have extremely low populations and low overall employment. Table 8 presents information about the distribution of manufacturing plants, manufacturing employment, and total employment. In 1947, the manufacturing belt accounted for 70 percent of large manufacturing plants and 70 percent of overall manufacturing employment across all sizes of plants. But it accounted for 49 percent of total employment, so the LQ is 1.42 for each measure of manufacturing activity. The Piedmont region accounted for 8.4 percent of large plants and 6.7 percent of overall manufacturing employment. The LQ for the Piedmont region is 1.53 with activity measured by the number of large manufacturing plants and 1.22 for activity measured by total manufacturing employment. These numbers indicate that manufacturing employment in the Piedmont area in 1947 tended to be in larger plants compared with the nation as a whole. In contrast, the California region had a smaller share in large plants than its share of the total employment, and so it has an LQ of less than one.

Figure 3 presents the analogous map of the location of large manufacturing plants (250 or more employees) for 1999. For 1999 there are 12,586 plants, about 20 percent more than the 10,217 in 1947.[11] The similarities are impressive. Concentration is still heavy in the Megalopolis and around the Great Lakes. The Piedmont region in the South remains an area of high concentration, and Appalachia is still devoid of manufacturing.

But the contrast is also impressive. Manufacturing is clearly much more dispersed. There are only approximately 2,000 more plants, but it seems like many more because they are so much more spread out. The big stories here are states like Tennessee, which had little

[11] It is interesting to note that total employment is lower, so mean size conditional on being above 250 is lower.

manufacturing in 1947, but now has an extremely high manufacturing presence. Similar stories can be told for Arkansas and Mississippi. Prairie states like the Dakotas and Nebraska have also had substantial increases in manufacturing.

Table 8 presents the shares for the various regions. The decline in the manufacturing belt is striking. Notice that the manufacturing belt lost a third of its employment and a third of its large manufacturing establishments. This decline is masked in a comparison of Figures 2 and 3 as the spreading out of plants in the 1999 figure gives the illusion that there are more plants. It is notable that the manufacturing LQ for the manufacturing belt equals one in 1999. The term *manufacturing belt* is no longer appropriate for this area. Note that manufacturing remains concentrated in the Piedmont area. The LQs for this area, already bigger than one in 1947, increased further. Finally, note that the LQ for the residual region (in plants with 250 or more employees) rose from 0.45 in 1947 to more than 0.9 in 1999; the residual region now looks much more like the nation as a whole than it did in 1947.

We conclude that between 1947 and 1999, manufacturing left the manufacturing belt in a big way, areas with low manufacturing increased their level of manufacturing, and the southern area, which was already concentrated in 1947, further increased in concentration.

4 Specialization by Urbanization

The composition of economic activity varies substantially across big cities, small cities, and rural areas. It is obvious that certain types of economic activities in big cities (e.g., Broadway plays, stock exchanges) are very different from the activities in rural areas (e.g., dairy farms). But there are even differences between medium-sized cities and small cities. In his classic work on central place theory, Christaller (1966) shows how cities that vary in size differ systematically in the range and type of goods that can be found in the city (see also Berry (1967)). Another important reference that documents the strong specialization patterns with city size is Henderson (1988). This section follows in the tradition of this earlier work and lays out the specialization patterns by degree of urbanization in the United States and Canada at the turn of the 21st century.

4.1 A Measure of Urbanization

We begin by defining a county-based measure of urbanization used previously in Holmes (1999). For each county i and j, we determine the distance d_{ij} between the two in miles.[12] We define the neighborhood of county i to be

$$N_i = \{j| \text{ such that } d_{ij} \leq 30\}.$$

This neighborhood includes counties within 30 miles of county i, including county i itself. Define the *neighboring population of* i to be the total population of all the counties in N_i. We use population as determined by the 2000 census.

A more common method of classifying locations by city size is to use metropolitan statistical areas (MSAs) defined by the Census Bureau. The MSA classification system is convenient for many purposes, and we often use it in our own work. When we redo the analysis below with the MSA classification system, we obtain similar results. We choose to report the results with the neighboring population definition because we think it is preferable for the particular exercise we conduct. The MSA classification system treats all geographic points in a metropolitan area equally, even though they can be quite different. For example, the New York consolidated MSA includes high-density Manhattan as well as relatively low-density portions of northeastern Pennsylvania. In addition, definitional issues arise as to whether San Francisco and Oakland should be considered a part of the same MSA. Ciccone and Hall (1996) raise these issues and argue that population density is the appropriate way to measure concentration of people. However, county boundaries can vary in arbitrary ways, and these boundaries can have a big influence on a measure of population density. Furthermore, density at a particular point may not be all that is relevant; population nearby this point may also matter. An ideal measure for our purposes is one that takes each geographic point and counts the exact number of people within a fixed distance from that point. Our county-based measures fall short of this ideal because county boundaries are arbitrary—the neighborhood N_i defined above does not form a perfect circle. However, for the purposes at hand the approximation is a reasonable one as discussed further in Holmes (1999).

[12] We use the longitude and latitude coordinates for each county provided in the 2000 census. In most cases this is the geographic centroid, but in other cases it is just some point lying within the county.

The variation in neighboring population is quite dramatic. Using the data from the 2000 census, the maximum value is 15.4 million and is attained in Hudson County, New Jersey. This is across the river from Manhattan (neighboring population 15.0 million). The lowest value, 67, is obtained for Loving County, Texas. Weighting by population, the median value is 1.2 million. Thus, half of the U.S. population lives in a county with neighboring population of 1.2 million or greater.

Our interest is in how city specialization varies by city size. As before, we use employment to measure economic activity. We divide the total U.S. employment of 165 million into quartiles based on the neighboring population of the location of employment. This partitions the United States into four parts each with 25 percent of total U.S. employment.[13] We henceforth refer to these groups as *urbanization quartiles*.

Table 9 presents summary statistics for the quartiles. While each quartile has 25 percent of total U.S. employment, the quartiles differ dramatically in land mass. The bottom quartile has approximately 87 percent of the U.S. land mass, and the top quartile has less than 1 percent. While the bottom quartile has a slightly higher share of population at 27.7 percent, the top quartile has a slightly lower share of population at 23.4 percent. This pattern reflects the fact that employment as a percentage of the population is higher in large cities than in rural areas, rising from 53 percent in the bottom quartile to 62.9 percent in the most urban quartile. In part, this difference is because the rural population is older and more likely to be retired. As shown in the table, 14 percent of the rural population is 65 or older while this is true for only 11 percent of the most urban population. The table also shows that individuals in urban areas tend to be much better educated than their counterparts in rural areas. From this we can anticipate that there will be relatively more white collar jobs in urban areas compared to rural areas.

[13] Lumpiness in county employment size prevents an exactly equal division. The percentages for the four groups are 25.0, 25.0, 24.6, and 25.4.

4.2 Specialization

Table 10 presents LQs for the major sectors in each of the four urbanization quartiles.[14] Since each quartile has 25 percent of all employment by definition, the sum of the location quotients across the four quartiles must add up to four. In particular, if all of an industry's employment is in a given urbanization quartile, then the LQ in that quartile will be four and the rest will be zero. The final column shows the difference between LQ_4 and LQ_1. We define this difference as a summary measure of an industry's tendency to locate in more urban locations. The table is sorted in terms of increasing urbanization. Table 11 presents industries defined at the narrowest level possible (six-digit NAICS). In this table we present the ten narrowly defined industries that are least urbanized as well as the ten most urbanized. For these examples the degree of specialization is extreme. The six-digit industry iron ore mining has an LQ_1 of approximately four, meaning that virtually the entire industry is in the least urbanized quartile. At the other extreme the six-digit commodities exchange industry has an LQ_4 of approximately four, meaning that virtually the entire industry is in the top quartile.

Turning back to Table 10 for the major sectors, notice that there is a gap in the urbanness measure between -0.36 and -0.22 and again between 0.20 and 0.58. We use these gaps to define three groups of sectors. The sectors with urbanness below -0.36 we call *rural sectors*. The sectors above 0.58 we call *urban sectors*. The sectors in the intermediate range between -0.22 and 0.20 are *diffuse sectors*.

Consider first the rural sectors. Observe that with one exception, the LQ monotonically declines across the urbanization quartiles. The one exception is mining, where the LQ increases from 0.28 to 0.44 from quartile 3 to quartile 4; in both quartiles the LQ is quite small. It is no surprise to see agricultural sectors and mining classified as rural sectors. It may be more of a surprise to see manufacturing in this group. In earlier days manufacturing activity was associated with cities. In the U.S. economy today, rural areas are net exporters

[14] This is the same set of sectors used in Section 3.1 with the addition of proprietorships as an additional sector of employment. This sector consists of self-employed individuals. Unfortunately, the data for proprietorships are not broken down by industry, or we would have allocated this employment to each of the various sectors.

of manufactured goods, while large cities are net importers. For this sector, the LQ is 1.18 and 1.13 in the bottom quartiles and 0.93 and 0.76 in the top quartiles. Thus, both of the bottom quartiles tend to specialize in manufacturing in contrast to the agricultural sectors, where just the very bottom quartile has an LQ above one. Also among the rural sectors are utilities and government. The utility sector is a small sector (0.4 percent of the employment) and consists primarily of power plants, which tend to locate disproportionately in rural areas. Government employment is a large sector of the economy. The fact that government concentrates in rural areas initially came as a surprise to us. We suspect this might have something to do with the absence of economies in rural areas for various government services such as education. Also, some government activity, such as military bases and national parks, tends to be in rural areas.

Next consider the diffuse sectors. These consist of the types of goods, such as retail and health care, that are generally regarded as nontraded goods. Even though the LQs for these sectors are fairly close to one in all urbanization quartiles, there are some patterns in the data. The LQ for retail is 1.08 in quartile 1 and monotonically declines to 0.86 in quartile 4. Analogously, the LQ for health care takes on its highest value of 1.04 in the most rural quartile. One factor contributing to this pattern is that less urbanized areas have a higher ratio of population to employment as noted above. (Retired people shop and require medical attention even if they do not work.) Furthermore, the inability to exploit economies of scale in rural areas also plays some role. (More employment is needed to provide a given service level.) Note that there is a slight tendency for proprietorship employment to be concentrated in rural areas, also presumably due to the absence of scale economies in rural areas, which tends to favor small-scale enterprises.

Finally, consider the urban sectors. Without exception, the LQ monotonically increases in urbanness for all of these sectors. This group of sectors includes service-type goods such as wholesale trade, finance, and business services. It also includes transportation, because major transportation hubs are concentrated in big cities. It is interesting that all of the new sectors of the economy added through NAICS, such as "Information and Professional Services," are urban industries.

In the top half of Table 12, we combine all the rural sectors and similarly all the diffuse and all the urban sectors. The total employment share of the rural industries is 0.26. By coincidence, the total employment of the urban sectors is also 0.26. The remaining diffuse group accounts for 0.48. The table reports the weighted average of the LQs for each group. For the rural sectors, on average the LQ falls from 1.33 in quartile 1 to 0.76 in quartile 4, a fall on the order of 50 percent. In the urban sectors, on average the LQ increases from 0.58 in quartile 1 to 1.38 in quartile 4, an increase of approximately 70 percent. These values indicate that for over half of the economy, there is substantial specialization by the degree of urbanization.

In the bottom half of Table 12, we conduct a similar exercise using narrow industries rather than broad sectors. We sorted narrow industries by the urbanization measure, and—to be comparable to what we did with the broad sectors—picked cutoffs to get approximately one quarter of employment classified in rural industries and one quarter in urban industries.[15] The quantitative results with narrow industry definitions are fairly close to what we got with the broad definitions at the sector level. For example, using the broad sectors, and looking at the top quarter of industries, we see that the LQ increases from 0.58 in the bottom quartile of locations to 1.38 in the top quartile. This change is similar to the increase from 0.50 to 1.46 we get with the narrowly defined industries. We conclude that the main pattern of specialization occurs at the broad sector level.

This result may come as a bit of a surprise, because we know that in *some* of the most narrowly defined sectors we see the highest specialization. But there is an offsetting factor as well. Consider, for example, the major sector "Finance and Insurance." Its index of urbanization is 0.71, making it one of the most urbanized sectors. Within this sector are the narrow industries "Commodity Contracts Dealing" and "Commodity Exchanges" that are extremely concentrated in the largest cities with indices of 3.04 and 3.90. But this sector also has industries such as "Consumer Lending" and "Commercial Banking" that are relatively spread out, having indices of 0.10 and 0.14. We conclude that within a broad industry there are some narrow industries that provide retail-like services that are spread out and other industries that are tradable and specialized.

So far we have ignored the dartboard issues highlighted by Ellison and Glaeser (1997).

[15] It is not exactly one quarter because of lumpiness in the industries.

These effects of chance are important when a small number of establishments are allocated across a large number of locations. Here we have only four locations, all of equal size. In the case where industries are defined at the sector level, all industries have a large number of establishments. If establishments in these sectors were randomly assigned to the locations with a 25 percent probability, the resulting location quotients would be close to one, as we have verified by simulation exercises. Thus, the specialization illustrated in Table 12 for major sectors has nothing to do with dartboard issues. In the case of narrowly defined industries, some industries have relatively few establishments, even as few as 10 or 11 in some instances. In such cases, the dartboard issues emphasized by Ellison and Glaeser (1997) have a greater potential to play a role. Nevertheless, we have done some simulation analysis at the level of narrow industries and found that these dartboard issues are not quantitatively significant for our case. This should not be much of a surprise since the median industry has over 1,000 establishments and these establishments are being allocated across only four locations. In such a situation we expect the law of large numbers to be in force, in which case the random factors average out.

4.3 Canadian Data

Having examined the U.S. data, we now turn to the Canadian data. We classify the 288 Canadian census divisions in the same way that we classified the 3,000 U.S. counties on the basis of neighboring population. (Here we use population from the 2001 census.)[16] From Table 9, the cutoffs for neighboring population that define the quartiles are used to define the quartiles for the United States. In the United States, the median county, weighted by total employment, has 1.4 million in neighboring population. In Canada, the median census division has 1 million in neighboring population, again weighted by employment. The 75th percentile is 3.5 million for the United States and 3.5 million for Canada. Differences emerge at the 25th percentile; 0.4 million for the United States and 0.2 million for Canada.

Table 13 reports the LQs for Canada, sorting industries by the urbanization index $LQ4$–$LQ1$, as before. For the most part, the pattern of location for the Canadian industries is

[16] For locations in Canada near the U.S. border, we count as neighbors only Canadian locations. This is analogous to what we did when calculating neighboring population in the United States.

very similar to the pattern for the United States. With two exceptions, the urban sectors in the United States are urban sectors in Canada. The education sector is not an urban sector in Canada, but that is due to a discrepancy in data collection. In Canada, education includes public schools, and employment is disproportionately high in rural areas because of the lack of economies of scale. In the United States such employment is treated as government employment. So the education sector in the United States is largely university employment, which is concentrated in big cities. The other exception is "Transportation and Warehousing."

Turning to the industries that were classified as rural in the United States, the major exception in Canada is manufacturing. This is an urban industry in Canada. In Canada, the manufacturing base has been maintained in major urban areas (e.g., Toronto and Montreal) in a way that is different from in the United States.

We can also compare the United States and Canada for more detailed industry definitions. The highest level of comparable detail is the four-digit NAICS. At this level of detail, there are approximately 300 industries.[17] We calculated the LQ for each urbanization quartile for each industry for both Canada and the United States. The construction sector and the wholesaling sectors are apparently not harmonized yet so we can only directly link up 256 industries. For these industries, we compared the urbanization measure (i.e., LQ4–LQ1) in the two countries and found a correlation of 0.65. This indicates a strong tendency for the pattern of concentration to be the same in the two countries.

5 Some Facts About Specialized Industries

The preceding two sections focused on characterizing the extent to which various industries are concentrated. In this section we present selected facts about concentrated industries. We make two kinds of comparisons. First, we look *across* industries. We compare the characteristics of industries that have high geographic concentration with those that have

[17]The Canadian Business Patterns data include industries such as crop farming and livestock farming that are not in the U.S. County Business Patterns data. There are 321 four-digit NAICS industries in the Canadian Business Patterns and 290 in the U.S. CBP.

low concentration. Second, we look *within* industries. Here we compare establishments in the same industry across locations. We compare these plants in areas of concentration with those plants in the same industry in isolated locations.

5.1 Across Industry Comparisons

In this subsection we discuss two industry characteristics that have been found to be associated with high industry concentration. The two characteristics are *establishment size* and *materials intensity*.. Using two-digit level manufacturing data for the United States over the period 1880–1987, Kim (1995) reports a positive correlation between geographic concentration and both establishment size and materials intensity. Table 14 uses the EG indices constructed in Section 3 to illustrate these tendencies.

The first set of figures reported in Panel A are calculated with the set of all 1,082 NAICS six-digit level industries from the 1999 CBP. Each industry is classified by average establishment employment size. The 1,082 industries are divided into four categories based on employment size so that the number of industries in each category is close to being the same as is feasible. The table reports that the average employment size in the bottom quartile of industries is 6.6 employees per establishment, and this rises to 167.2 in the top quartile of industries. The table also reports the mean EG index among the industries in each quartile. The mean rises from 0.027 in the bottom quartile to 0.062 in the top quartile, a sharp increase. This pattern has a fairly straightforward explanation. Industries with large establishments tend to be manufacturing industries while industries with small establishments tend to be retail and services. The industry with the largest average size (2,400 employees), "Guided Missile Manufacturing," has a γ of 0.13. The industry with the smallest average size (2.4 employees), "Footwear Repair," has a γ of 0.005. As discussed in Section 3, manufacturing industries tend to have large EG indices, while the EG indices are close to zero for retail and services. This result reflects the obvious fact that manufacturing industries tend to be industries with trade (and, hence, specialization) compared to retail and services.

The second set of figures in Panel A restricts attention to the 473 different six-digit NAICS manufacturing industries and defines quartiles in this subset as before. The average concentration index increases with average establishment size, just as it does when we consider the entire cross-section of all industries. The manufacturing industry with the smallest average size (5.3 employees), "Dental Laboratories," has a γ of 0.002. The next two on the list are "Retail Baking" ($\gamma = 0.007$) and "Quick Printing" ($\gamma = 0.001$). These manufacturing industries with very small establishment sizes are obviously "retail-like" in their nature, and so it is not surprising that they tend to be geographically dispersed.

We next consider *materials intensity,* which, following Kim (1995) and Holmes (1999, p. 314), we define to be purchased inputs as a percentage of total sales. This information is available only for manufacturing.[18] Panel B classifies manufacturing industries into quartiles by the materials ratio. The average value of the concentration index increases from 0.027 in the bottom quartile to 0.065 in the top. Industries with very high materials ratios tend to be first-stage processors of raw materials located near the source of these raw materials (and, hence, are geographically concentrated). For example, the industry with the highest share (0.88), "Soybean Processing," has $\gamma = 0.07$.

It is possible to make other cross-section comparisons. For example, if we classify the 1,082 six-digit NAICS industries in the CBP data into quartiles by average pay per employee, the resulting average γ increases from 0.015 in the bottom quartile to 0.060 in the top quartile. This result is not surprising, since manufacturing tends to be high-pay and concentrated, while retail is low-pay and dispersed. Interestingly, if we construct quartiles with just manufacturing, we get something of a U-shaped relationship with the quartiles defined by pay. At the top of the pay scale are industries like Petrochemical Manufacturing and Guided Missile Manufacturing that tend to be highly concentrated (γ equal to 0.38 and 0.13). But at the bottom of the pay scale are textile industries, and these also tend to be geographically concentrated.

5.2 Within Industry Comparisons

Within industries, establishment size, plant turnover, productivity, and plant specialization all vary across locations. These variations are related to the degree of local industry concentration. Holmes and Stevens (2002) find a positive correlation between an establishment's

[18] The data we use are from the 1997 Census of Manufactures.

size and local industry concentration. This correlation persists in virtually every major sector of the economy and for various degrees of industry and geographic aggregation. Importantly, it remains even after controlling for an establishments's own effect on local concentration. In the manufacturing sector, for example, "plants in the highest quintile of the own-plant excluded measure of concentration are on average 64% larger than plants in the lowest quintile" (Holmes and Stevens (2002), p. 683).

Dumais et al. (2002) examine the relationship between industry concentration and establishment turnover. Using plant-level data from the Census Bureau's Longitudinal Research Database, they find that plant births tend to disperse employment away from concentrated areas, whereas plant closures tend to reinforce existing concentrations. They also report that even though aggregate levels of concentration have remained constant, over time there is some shift in the centers of concentration. They suggest that small, young plants with high turnover that exist outside centers of concentration may serve as experimental draws in a search process for new centers of concentration.

There is a long-standing debate on the relative importance on productivity of localization (having many own-industry neighbors) versus urbanization (having many neighbors from any industry).[19] Early work, such as Sveikauskas (1975) and Segal (1976), reports evidence on the importance of urbanization. Later, Henderson (1986) reports evidence on the importance of localization. Using Census micro data, Henderson (2003) expands this work in a number of directions. In particular, he permits the relationship between productivity and city and industry size to differ by plant type (single plant firms or corporate plant), considers alternative sources of spillovers (employment, number of plants, and plant births), and introduces a time dimension to allow for what he calls "dynamic externalities." From this rich structure he finds that manufacturing externalities are very local and depend on the number of plants in the same industry and same county. For static externalities, this result holds for both single plant firms and corporate plants, while for dynamic externalities it seems to matter more for single plant firms. Another approach is to look at density of employment rather than levels. Ciccone and Hall (1996, p. 54) find that "a doubling of employment density

[19] External effects may also operate at the worker level. For example, Glaeser and Maré (2001) find that cities seem to speed the accumulation of human capital.

increases average labor productivity by around 6 percent."

One potential benefit of agglomeration is that the greater concentration of similar firms may result in a local network of intermediate goods providers and more specialization among firms. Evidence from Holmes (1999) shows that purchased inputs as a percentage of the value of output is greater in locations where local industry concentration is higher: "a plant with anywhere from 10,000 to 25,000 in own-industry neighboring employment has a purchased-input intensity that, on average, is three percentage points higher than a plant with fewer than 500 in own-industry neighboring employment" (Holmes (1999), p. 314). Ono (2001) also reports a connection between outsourcing and concentration.

6 Understanding the Location of Industry

This chapter has focused on presenting basic facts about patterns of industry location; it leaves to other chapters the job of discussing theories of firm location and their empirical importance. Nonetheless, we think it is useful to include a section that reports some observations related to the issue of *why* industries concentrate. In this discussion, we focus on three factors that affect the location of industry: natural advantage, what we call the *concentrated market factor*, and government policy. We recognize that there are other factors at work, but again we are relying on other chapters for a more comprehensive discussion of these factors.

6.1 Natural Advantage

Locations differ in innate physical attributes, such as climate, soil, minerals, and access to deep water ports. Differences in such natural advantages can lead to specialization. This follows from the standard analysis of Ricardian comparative advantage.

The natural advantage force is obviously of overwhelming importance in agriculture, an industry with a substantial degree of specialization. Figure 4 presents a dot density map for selected crops. (Each dot represents 50,000 tons of a particular crop.) Different colored dots represent different crops. Observe that for most of the crops, the colored dots demarcate areas with sharp boundaries. For example, the eastern part of Arkansas has heavy rice production; the western part has essentially zero.

While natural advantage is, of course, the primary factor leading to specialization, we think it is worth noting that scale economies may be playing a supporting role. The types of scale economies we have in mind include the usual ones that lead to industry agglomeration. Different crops have various specialized intermediate inputs, and, of course, transportation costs are important, so what we refer to below as the *concentrated market factor* should matter. We expect knowledge spillovers to matter as well. There is also the issue that custom hybrid seeds are developed for particular areas to exploit the climate; this is a fixed cost. (See Griliches (1957) for a classic study of the diffusion of hybrid corn.)

These issues can be illustrated nicely with the example of sugar beets. Hoover (1948, p. 33) presents a map of the joint location of sugar beet farms and processing plants as of 1939. Figure 5 uses recent data to update this earlier map, though the pattern has changed little. The largest area of beet production is in the Minnesota/North Dakota region, but there are several other regions of production. The locations of the 39 sugar beet processing plants (from the 1999 CBP data) are closely connected to the locations of beet production.

In the traditional classification of industries developed by Weber, some are *transportation-cost oriented* and others are *production-cost oriented*. Of the transportation-cost oriented, some are materials-oriented and others are market-oriented. The sugar beet processing industry is a classic example of a *materials-oriented* industry. The sugar content of sugar beets is roughly 15 percent by weight, so seven tons of beets must be used to get one ton sugar. The incentives to place processing near field crops and vice versa are obvious. The EG index for the processing industry is 0.13, a very high number. It is clear that this high concentration of processing is being driven by the high geographic concentration of the upstream crops.

Sugar beet processing plants are relatively large. In 1999, the average plant size was 186 employees, more than four times the average plant size of 46 employees for manufacturing as a whole. This large plant size suggests that scale economies may be important. Note that there are several production regions on the map with access to only a single plant. This is true in southern California and in the eastern Montana/western North Dakota region. If scale economies in processing are important and if local beet farming does not reach a large enough scale to support the processing plant, the local beet farming activity may have to shut down. The industry's trade literature suggests that the issue of economies of scale is crucial. One reads about plants shutting down because they cannot contract for a sufficiently large number of acres. One reads about farmers not planting crops because a plant has closed.

This last point can be cleanly made with an example about a 4-H club in California.[20] (A 4-H club is a youth group that focuses, in part, on farming activities.) For 33 years this particular club had run a project in which the youth grew their own sugar beets and had them processed by the local plant. But no more. According to the article, "Imperial Sugar Co. decided to cease sugar production at its Woodland and Tracy processing plants by the end of the year.... This impacts the Glenn County 4-H sugar beet project since it is too far to transport beets to other facilities." The 4-H club would like to continue the production and in that sense has a "natural advantage." But this small production will not keep a 200-employee factory busy, and so it does not happen.

The role of comparative advantage in contributing to regional specialization in the United States does not receive much research attention, in contrast to the trade literature where there is substantial research on the importance of both Ricardian comparative advantage and Hecksher-Ohlin factors. A few exceptions are Ellison and Glaeser (1999), Kim (1995), and Hanson and Slaughter (2002). Ellison and Glaeser attempted to control for natural advantage characteristics by, for example, controlling for the amount of nearby agricultural production. They found that the mean EG index declined by 20 percent after doing so, and they conjecture it might fall by half if they had better data. We think this is an interesting line of work that would be fruitful if pushed further. But we want to introduce a word of caution that even for agricultural industries, scale economies may be playing some role in determining location patterns. To the extent this is true, the EG procedure will overstate the importance of natural advantage.

[20] See "4-H project ends with closure of sugar beet processing plant," by Christine Souza, January 10, 2001, in *Crisis on the Farm: A Special Series from Ag Alert,* California Farm Bureau Federation, www.cfbf.com/agalert/2001/crisisstory3.htm.

6.2 Concentrated Market Factor

When scale economies and transportation costs both matter, there is an incentive to *concentrate* a market. With a large market in the same place, economies of scale can be enjoyed and transportation costs avoided. In the literature this factor is also called the *variety effect*, or the *Chamberlinian* factor (see Fujita and Thisse (2002)), or the *intermediate-inputs factor*. The last term seems limiting because the logic applies not just to intermediate inputs but also to final goods, as will be the case in the example we consider.

There has been a substantial amount of theoretical work on this issue (Fujita et al. (1999), Fujita and Thisse (2002)), but little empirical work. How much of this factor can account for the agglomeration of industry is a key question. Papers on this topic are starting to appear and include Holmes (2002) and papers cited in the chapter by Rosenthal and Strange in this volume.

In this subsection we illustrate the empirical importance of this factor by considering the example of the beverage industry. This is a classic example of a weight-gaining industry—water is added to other materials in the final stage of production. Here there is an incentive to place production near consumers. But there are economies of scale, so we don't expect to see production in every random rural hamlet.

We narrow things further by looking at the beer industry and, in particular, the plant location patterns of the largest brewer, Anheuser-Busch. Anheuser-Busch has 12 breweries, and these plants are quite large; the largest in St. Louis has over 1,500 employees. With scale economies and high transportation costs, the concentrated market factor should matter here if it matters anywhere.

Figure 6 plots the location of this company's 12 breweries. The location of the breweries is from the TRI database (discussed in the Appendix), which provides the exact geographic coordinates of each factory. The black dots are the breweries; each red dot represents 10,000 people. There are two things to note about this graph. First, there are no plants in the extreme rural areas; i.e., plants tend to be close to population centers. Second, the plants appear to be spread out in a way that would be consistent with keeping transportation costs low. For example, we don't see two plants in the same metro area. We now provide some more formal evidence of these two propositions.

For evidence that plants are not in rural areas, consider the last column of Table 15. This column reports for each plant the neighboring population percentile for the county where the plant is located. As before, we use a distance of 30 miles to define neighboring population. For example, the percentile for the Fairfield, California, plant is 55.1. This means that the neighboring population of the plant location is higher than 55.1 percent of other locations in the continental United States (weighting the other locations by population). The lowest percentile across the 12 plants is 28.2 percent. If the plants were randomly distributed on the basis of population, the probability that all plants would land in the 28.2 percentile or higher is only $(1 - 0.282)^{12} = 0.019$. Thus, plants tend to be located away from rural areas in a disproportionate sense. This is consistent with theories from the concentrated market literature, which find that plants will tend to be located in areas with large home markets.

We now turn to the second claim that plants are spread out to minimize transportation cost. Note that two plants located in the same metro area would duplicate fixed costs but contribute nothing to reducing transportation costs. To get some idea of the extent to which the selected locations minimize transportation costs, we consider the following calculation. Assume for this analysis that Anheuser-Busch's sales at each location are proportionate to the population at each location. (Here we use the county as the location unit.) For the purpose of discussion, suppose that Anheuser-Busch delivers a case of Budweiser to every person in the United States. Assume that each location is supplied by the closest brewery. Table 15 presents the market share of each of the 12 breweries based on this calculation. We sort the breweries going from west to east. We also report the average distance per consumer served per brewery. The weighted average transportation distance across all the breweries is 178. Put another way, the average person in the United States is 178 miles from the closest brewery.[21]

As a benchmark, we first compare this cost to what the cost would be, in expectation, if plants were allocated randomly, like in Ellison and Glaeser (1997). Suppose the 12 plants are randomly distributed across counties in proportion to population, each dartboard toss

[21] This calculation excludes Alaska and Hawaii, but the populations of these states are quite small, so including them would have little effect.

being independent and identically distributed. After throwing 12 darts, we then calculate the average distance. We repeat this 10,000 times. We find that the average distance across simulations is 246 miles and that the standard deviation is 50 miles. In a fraction of only 0.004 of the simulations is the average distance less than or equal to the average actual distance of 178 miles. We can therefore strongly reject this dartboard model. Observed transportation cost is much lower than what would happen with a random distribution of plants.

For the sake of argument, one might try to make a case that the plant location choices reflect some systematic natural advantage factor. For example, suppose (hypothetically) that beer produced in rural areas were to taste bad compared to beer produced near urban areas. To give this alternative explanation a chance, we have also considered various randomizations like those above that exclude rural areas. It turns out that this exclusion led to virtually no reduction in expected cost of a randomly drawn configuration. Thus, even when rural locations are excluded, the random configuration does much worse than the actual configuration. The proposed natural advantage theory can explain why plants are not in rural areas but cannot account for the configuration of plants across nonrural areas.

We conclude that the evidence suggests that the concentrated market factor is an important determinant of the plant location strategy of Anheuser-Busch. Obviously, it is not the only factor. We believe an important area of future research is to disentangle the importance of the concentrated market factor from other factors. We believe that work with micro data sets that consider multi-plant firm location problems, as discussed here and in Holmes (2002), will prove useful in getting to the bottom of this issue.

6.3 Government Policy

One of the most important and controversial issues in location analysis is the role that policy plays in influencing location decisions. It is important because policies are something societies can change and economic analysis can provide guidance for these public choices. It is controversial because policy effects are difficult to pin down. Not only do researchers have to disentangle policy effects from other factors, they also have to deal with complicated causality issues.

It is impossible in this brief section to survey this literature. Instead we make mention of a few contributions that illustrate recent progress on this issue. For references to the earlier literature, see Herzog and Schlottman (1991).

One issue that has attracted much attention is the effect of environmental regulations on the location of economic activity. While earlier research often found relatively small effects, the more recent research of Becker and Henderson (2000), Greenstone (2002), and Henderson (1996) finds relatively large impacts. A hallmark of this new research is the use of richer data compared with previous studies, making it possible to do a better job of measuring the incidence of policy and of controlling for confounding factors. While the earlier literature generally treated the regulatory policies as occurring at the state level, this new literature points out that much policy is administered at the county level. In particular, because of federal law, each county is classified as being either in attainment or not in attainment and various restrictions are imposed in non-attainment counties. All three studies find that non-attainment status in a county leads to shifts in the location of polluting industries away from such counties.

A second issue that has attracted attention is the role that policy has played in regional shifts in manufacturing. As discussed in Cobb (1993), southern states have actively pursued a host of pro-business policies to attract manufacturing. These include policies hostile to labor unions like right-to-work laws as well as low taxes and subsidies. Progress in quantifying the effects of such pro-business policies has been hampered by the difficulty of distinguishing the effects of state policies from the effects of other state characteristics that are unrelated to policy. In a recent paper, Holmes (1998) uses a regression discontinuity design to try to control for these other factors. Specifically, he examines the differences in manufacturing levels and growth rates at state borders between "pro-business" and "anti-business" states. Near the border, other differences unrelated to policy are likely to be very similar, and so the difference at the border identifies the effect of policy. Holmes finds relatively large effects of policy near these borders.

We illustrate the idea by looking at an example industry in which the policy effects are so overwhelming that the point is crystal clear. Figure 7 is a map of county-level LQs for the

"Consumer Lending" industry NAICS 522291.[22] This industry is defined as "establishments primarily engaged in making unsecured cash loans to consumers." Included in this industry are establishments that make short-term loans with exorbitant interest rates. It is readily apparent that there is wide variation across the United States in the extent of this activity. The key thing to observe in this table is the sharp differences in state boundaries. Note in particular the state in the center close to the bottom that is very light but is surrounded on the left bottom and center by states that are completely dark. The state in question is Arkansas, and the surrounding states are Oklahoma, Louisiana, and Mississippi. The sharp difference at the border makes it obvious that there must be a policy difference between Arkansas and its neighboring states. One policy difference is the unusually strong anti-usury laws that exist in Arkansas and the absence of those laws in the neighboring states.[23]

Observe that in the region of southern states like Oklahoma, Texas, Louisiana, and Mississippi, activity in the consumer lending industry tends to be high. If we worked with just state-level data, we might not be able to distinguish a policy explanation from an argument that the South has a natural advantage in this industry. Indeed, it is obvious that if we regress state-level LQs on latitude we will get some explanatory power. It is only by looking at detailed geographic data near the border that we get a "smoking gun" effect of policy.

7 Conclusion

In this chapter, we reviewed data sources that are useful for studying the geographic specialization of industry in the United States and Canada. In particular, we emphasized the use of detailed data on establishment counts by employment-size categories and fine industry and geographic classifications. Basic facts about specialization were elicited from these data using standard measures, such as employment location quotients.

The chapter also touched on a classic, unresolved question: Why do specialization patterns emerge as they do? In recent years, a resurgence of work in economic geography has

[22] The LQs are based on County Business Patterns 2000 employment.

[23] See "The bad side of town," in the Economist, 11/28/98, pp. 30–31.

made theoretical inroads into this interesting and important question. However, many open issues remain, and empirical work continues to lag behind the theoretical frontier. Progress has been hampered by difficult identification problems, such as how to sort out the relative importance of natural advantage versus scale economy explanations for industry agglomerations (as explained by Ellison and Glaeser (1999)). We are optimistic that strategies can be devised to overcome these hurdles (e.g., by looking at multi-location firms as in Holmes (2002) and in the Anheuser-Busch example of Section 7) and that much progress will be made in the coming years.

Acknowledgements

Holmes acknowledges support from the NSF through grant SES 0136842. The authors thank Edward Glaeser, Vernon Henderson, and Jacques-François Thisse for helpful comments. They also thank Alice Schoonbroodt and Brian Rowe for research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Board of Governors of the Federal Reserve System, or the Federal Reserve System.

Appendix

This Appendix provides details about the various data sources discussed in the text.

Sources of the CBP Data

The CBP data are available in computer readable files from 1974 to the present. The years 1974–1976 are distributed, for a small fee, on a custom-made CD-ROM by the U.S. National Archives and Records Administration (NARA at www.nara.gov); NARA does not distribute data on the Web. Beginning in 1977, the CBP data are archived at the Inter-university Consortium for Political and Social Research (ICPSR at www.icpsr.umich.edu) and are freely available on the Web to participating institutions; for example, the data for 1977 are in ICPSR file 8464. Beginning in 1986, the data are also distributed on CD-ROM by the Census Bureau. The Census Bureau also tabulates these data on a ZIP code basis in a publication called *ZIP Code Business Patterns* that is available on CD-ROM with data beginning in 1994.

Additional information about the CBP data and an establishment-level version of the 2000 CBP are available for download at www.econ.umn.edu/~holmes/data/CBP.

Mean Employment by Size Class

Table 1 in the text reports average establishment employment in each establishment-size class based on the CBP data. For establishment-size classes with fewer than 1,000 employees, this average is calculated by taking the total employment in the United States in the size class and dividing by the total number of establishments in the size class. Unfortunately, the aggregate data for the United States do not break down the 1,000-plus category into the four narrow size classes for which we have establishment count data. Therefore, we estimate the mean sizes for these four categories using the following estimation procedure. We assume that a log normal distribution of employment holds for establishments in the 1,000-plus and up class. We use a generalized method-of-moments to estimate the parameters of this distribution, using the following as moments: (1) total employment in the 1,000-plus class and (2) establishment counts in the four component narrow size classes. The reported cell averages are for the estimated distribution.

Economic Census Data

As discussed in the text, the Economic Census also releases establishment cell counts at the county level. The data on establishment counts with 250 or more employees used to construct Figure 2 are from the 1947 Census of Manufactures. These data were published in the *1949 County and City Data Book* (variable #42) and are available in computer readable files, along with all the data in the county and city data books for the period 1944–1977, from ICPSR study 7735.

The Census of Manufactures has published cell counts by detailed industry and county and size class since at least 1954. This publication is called *Location of Manufacturing Plants*.

A computer file with this data is available for 1972 from NARA (www.nara.gov). The data for later Economic Census years are available from the Census Bureau on CD-ROM.

Canadian Business Patterns and Census Data

Section 3 uses province-level census data on total employment by NAICS sector for 2001; these data are published by Statistics Canada (2001) (www.statcan.ca) and are available on the Web. Two aggregates (shown in Table 4 in the text) were constructed from the eastern and western provinces to facilitate the regional comparisons.

Section 4 uses data from the Canadian Business Patterns CD-ROM. These data report cell counts at the four-digit NAICS level by Census Division (the analog of a U.S. county) and employment-size category. There is no information on total employment from which to calculate mean employment in the cell, so we used the U.S. average sizes as an estimate. The cells 1–4, 5–9, 10–19, 20–49, and 50–99 correspond to cells (or groups of cells) in the U.S. data and so we directly use the U.S. mean. The cells 100–199 and 200–499 are slightly different from their counterparts 100–249, 250–499, in the U.S. data. We adjust the estimates for Canada so that the mean relative to the range is the same as in the United States. For example, the mean employment in the U.S. data for the cell 100–249 is 150.1, which is a fraction 0.336 of the way between 100 and 249. We therefore set the cell average for the 100–199 range in Canada to 133.2, because this is a fraction 0.336 of the way between 100 and 199.

BEA Data

We obtained data from the U.S. Department of Commerce's Bureau of Economic Analysis (BEA at www.bea.gov) on government employment and proprietorships (self-employment). BEA data are available on CD-ROM and on the Web. The BEA aggregates certain counties that the Census Bureau treats separately; therefore, the BEA data have 3,110 distinct counties compared to the 3,131 counties that meet the Census Bureau's definition. For example, Virginia treats independent cities as counties and separates them from the surrounding area, while the BEA merges these cities with the surrounding areas. In Section 4 we use the BEA county definitions.

Other Data

We used establishment-level data from the Toxic Release Inventory (TRI) collected by the U.S. Environmental Protection Agency to identify the locations of the Anheuser-Busch plants. These data are available on the Web at www.epa.gov/tri/. These data provide the location (including exact geographic coordinates) of establishments required to file toxic release forms.

The data on agricultural crops were obtained from the Agricultural Statistics Data Base (www.nass.usda.gov) produced by the U.S. Department of Agriculture, National Agricultural Statistics Service.

References

Becker, R. and J.V. Henderson (2000), "Effects of air quality regulations on polluting industries", Journal of Political Economy 108:379–421.

Berry, B.J.L. (1967), Geography of Market Centers and Retail Distribution (Prentice-Hall, Engelwood Cliffs).

Christaller, W. (1966), Central Places in Southern Germany (Prentice-Hall, Englewood Cliffs), translated by C.W. Baskin.

Ciccone, A. and R.E. Hall (1996), "Productivity and the density of economic activity", American Economic Review 86:54–70.

Cobb, J.C. (1993), The Selling of the South: The Southern Crusade for Industrial Development, 1936–1990, 2nd edition (University of Illinois Press, Urbana, Illinois).

Combes, P.-P. and H. Overman (2004), "Spatial distribution of economic activities in the European Union", in: J.V. Henderson and J.-F Thisse, eds., Handbook of Regional and Urban Economics, vol IV (North Holland, Amsterdam).

Dinlersoz, E.M. (2002), “Cities and the organization of manufacturing”, Regional Science and Urban Economics, forthcoming.

Dumais, G., G. Ellison, and E.L. Glaeser (2002), “Geographic concentration as a dynamic process”, Review of Economics and Statistics 84:193–204.

Duranton, G. and H. Overman (2002), “Testing for localisation using micro-geographic data”, Centre for Economic Policy Research Discussion Paper No. 3379.

Duranton, G. and D. Puga (2004), “Micro-foundations of urban increasing returns: Theory”, in: J.V. Henderson and J.-F Thisse, eds., Handbook of Regional and Urban Economics, vol IV (North Holland, Amsterdam).

Ellison, G. and E.L. Glaeser (1997), “Geographic concentration in U.S. manufacturing industries: A dartboard approach”, Journal of Political Economy 105:889–927.

Ellison, G. and E.L. Glaeser (1999), “The geographic concentration of industry: Does natural advantage explain agglomeration?” American Economic Review 89:311–316.

Evans, D.S. (1987), “Tests of alternative theories of firm growth”, Journal of Political Economy 95:657–674.

Fritz, W.G. (1943), “Natural resource–minerals”, in: Industrial Location and Natural Resources (U.S. Government Printing Office, Washington, D.C).

Fujita, M., Y. Kanemoto, and J.V. Henderson (2004), “Spatial distribution of economic activities in East Asia”, in: J.V. Henderson and J.-F Thisse, eds., Handbook of Regional and Urban Economics, vol IV (North Holland, Amsterdam).

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

Fujita, M. and J.-F. Thisse (2002), Economics of Agglomeration: Cities, Industrial Location, and Regional Growth (Cambridge University Press, New York).

Gabaix, X. and Y. Ioannides (2004), "Evolution of city size distribution", in: J.V. Henderson and J.-F Thisse, eds., Handbook of Regional and Urban Economics, vol IV (North Holland, Amsterdam).

Gardocki, Jr., B.C. and J. Baj (1985), "Methodology for estimating nondisclosure in county business patterns", Center for Governmental Studies, Northern Illinois University.

Glaeser, E.L., H.D. Kallal, J.A. Scheinkman, and A. Shleifer (1992), "Growth in cities", Journal of Political Economy 100:1126–1152.

Glaeser, E.L. and D.C. Maré (2001), "Cities and skills", Journal of Labor Economics 19:316–342.

Greenstone, M. (2002), "The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 Clean Air Act admendments and the Census of Manufactures", Journal of Political Economy 110:1175–1219.

Griliches, Z. (1957), "Hybrid corn: An exploration in the economics of technological change", Econometrica 25:501–522.

Hanson, G.H. and M.J. Slaughter (2002), "Labor-market adjustment in open economies: Evidence from U.S. states", Journal of International Economics 57:3–29.

Henderson, J.V. (1986), "Efficiency of resource use and city size", Journal of Urban Economics 19:47–70.

Henderson, J.V. (1988), Urban Development: Theory, Fact, and Ilusion (Oxford University Press, New York).

Henderson, J.V. (1996), "Effects of air quality regulation", American Economic Review 86:789–813.

Henderson, J.V. (2003), "Marshall's scale economies", Journal of Urban Economics 53:1–28.

Herzog, Jr., H.W. and A.M. Schlottman (1991), Industrial Location and Public Policy (University of Tennessee Press, Knoxville, Tennessee).

Holmes, T.J. (1998), “The effects of state policies on the location of industry: Evidence from state borders”, Journal of Political Economy 106:667–705.

Holmes, T.J. (1999), “Localization of industry and vertical disintegration”, Review of Economics and Statistics 81:314–325.

Holmes, T.J. (2002), “The role of cities: Evidence from the placement of sales offices”, Federal Reserve Bank of Mineapolis Staff Report 298.

Holmes, T.J. and J.J. Stevens (2002), “Geographic concentration and establishment scale”, Review of Economics and Statistics 84:682–690.

Hoover, E.M. (1948), The Location of Economic Activity (McGraw Hill, New York).

Kim, S. (1995), “Expansion of markets and the geographic distribution of economic activities: The trends in U.S. regional manufacturing structure, 1860-1987”, Quarterly Journal of Economics 110:881–908.

MacDonald, J.M. (1985), “Dun and Bradstreet business microdata research applications, and the detection and correction of errors”, Journal of Economic and Social Measurement 13:173–185.

Maurel, F. and B. Sédillot (1999), “A measure of the geographic concentration of French manufacturing industries”, Regional Science and Urban Economics 29:575–604.

Ono, Y. (2001), “Outsourcing business service and the scope of local markets”, U.S. Census Bureau, Center for Economic Studies, Working Paper 01-15.

Pred, A. (1965), “The concentration of high-value-added manufacturing”, Economic Geography 41:108–132.

Rosenthal, S.S. and W.C. Strange (2004), “Micro-foundations of urban increasing returns: Empirical evidence”, in: J.V. Henderson and J.-F Thisse, eds., Handbook of Regional and Urban Economics, vol IV (North Holland, Amsterdam).

Rosenthal, S.S. and W.C. Strange (2001), “The determinants of agglomeration”, Journal of Urban Economics 50:191–229.

Segal, D. (1976), "Are there returns to scale in city size?" Review of Economics and Statistics 58:339–350.

Statistics Canada (2001), Canadian Business Patterns, CD-ROM.

Sveikauskas, L.A. (1975), "The productivity of cities", Quarterly Journal of Economics 89:393–413.

U.S. Dept. of Commerce, Bureau of the Census (1978), County and City Data Book [United States] consolidated file: City Data, 1944–1977 [computer file].Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census [producer], 1978. Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research [distributor], 2000.

U.S. Department of Commerce, Bureau of the Census (1985), County Business Patterns, 1977 [United States]: U.S. Summary, State, and County Data [computer file]. Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census [producer], 1985. Ann Arbor, Michigan: Interuniversity Consortium for Political and Social Research [distributor], 1986.

U.S. Department of Commerce, Bureau of the Census (2002a), County Business Patterns 2000, United States. Washington, D.C.

U.S. Department of Commerce, Bureau of the Census (2002b), County Business Patterns 1999–2000. CD-ROM C1-E00-CBPX-01-US1.

U.S. Department of Commerce, Bureau of Economic Analysis (2002), Regional Economic Information System (REIS), 1969–2000. CD-ROM RCN-0295.

Table 1: Mean Employment by Size Category
2000 County Business Patterns

Employment Size Range	Average Employment	Number of Establishments
1-4	1.7	3,807,810
5-9	6.6	1,365,401
10-19	13.5	895,312
20-49	30.2	613,540
50-99	68.8	212,293
100-249	150.1	123,924
250-499	340.7	32,691
500-999	681.3	12,170
1,000-1,499	1208.8	3,170
1,500-2,499	1892.9	2,188
2,500-4,999	3374.7	1,138
5,000 or more	9592.0	411

Source: Mean employment estimates for categories above 1,000 use authors' algorithm discussed in the Appendix.

Table 2: Comparing Actual and Estimated CBP Data

NAICS	Industry	Gini actual data	Gini estimated data
233	Building, developing & general contracting	0.119	0.115
234	Heavy construction	0.241	0.247
235	Special trade contractors	0.097	0.096
311	Food mfg	0.279	0.284
323	Printing & related support activities	0.198	0.198
327	Nonmetallic mineral product mfg	0.191	0.192
337	Furniture & related product mfg	0.348	0.346
339	Miscellaneous mfg	0.201	0.201
421	Wholesale trade, durable goods	0.085	0.085
422	Wholesale trade, nondurable goods	0.114	0.114
441	Motor vehicle & parts dealers	0.087	0.087
442	Furniture & home furnishing stores	0.073	0.076
443	Electronics & appliance stores	0.088	0.092
444	Bldg material & garden equip & supp dealers	0.091	0.088
445	Food & beverage stores	0.073	0.073
446	Health & personal care stores	0.091	0.091
447	Gasoline stations	0.191	0.193
448	Clothing & clothing accessories stores	0.089	0.089
451	Sporting goods, hobby, book & music stores	0.097	0.096
452	General merchandise stores	0.121	0.122
453	Miscellaneous store retailers	0.066	0.066
454	Nonstore retailers	0.163	0.158
484	Truck transportation	0.201	0.199
488	Transportation support activities	0.248	0.252
541	Professional, scientific & technical services	0.163	0.157
551	Management of companies & enterprises	0.173	0.177
611	Educational services	0.226	0.211
621	Ambulatory health care services	0.075	0.074
721	Accommodation	0.302	0.304
722	Food services & drinking places	0.060	0.060
811	Repair & maintenance	0.077	0.076
812	Personal & laundry services	0.057	0.055
813	Religious, grantmaking, civic, prof & like organizations	0.094	0.094

Source: Authors' calculations using the 2000 County Business Patterns.

Table 3: Employment Location Quotients for the United States

	Share of National Employment (percent)	West South Central	East North Central	East South Central	Middle Atlantic	Mountain	New England	Pacific	South Atlantic	West North Central
Crop and Animal Production	2.38	1.47	0.87	1.83	0.38	0.99	0.28	0.96	0.71	2.35
Agriculture, Forestry, Fishing and Hunting excluding Crop and Animal Production	0.15	1.13	0.41	1.78	0.34	0.90	0.69	2.09	1.18	0.50
Mining	0.39	3.66	0.42	1.28	0.34	2.74	0.12	0.50	0.60	0.69
Utilities	0.52	1.17	1.04	0.95	1.09	1.12	1.04	0.72	0.99	1.01
Construction	4.57	1.09	0.89	1.00	0.81	1.44	0.81	1.02	1.13	0.90
Manufacturing	12.88	0.86	1.36	1.34	0.85	0.65	1.06	0.94	0.88	1.03
Wholesale Trade	4.56	0.97	1.00	0.88	1.14	0.88	0.94	1.14	0.89	1.00
Retail Trade	10.97	0.99	1.01	1.05	0.94	1.04	1.01	0.92	1.06	1.01
Transportation and Warehousing	2.68	1.12	0.98	1.06	1.07	1.03	0.68	1.05	0.90	1.02
Information	2.41	0.89	0.79	0.67	1.16	1.03	1.18	1.26	0.95	1.04
Finance and Insurance	4.46	0.84	1.00	0.75	1.34	0.88	1.35	0.92	0.91	1.02
Real Estate and Rental and Leasing	1.43	1.00	0.85	0.73	1.04	1.14	0.86	1.27	1.07	0.77
Professional, Scientific, and Technical Services	4.73	0.88	0.86	0.62	1.19	1.01	1.06	1.21	1.10	0.75
Management of Companies and Enterprises	2.07	1.01	1.17	0.82	1.21	0.74	0.99	0.94	0.86	1.03
Administrative and Support and Waste Management and Remediation Services	6.01	1.11	0.92	0.81	0.90	1.09	0.88	1.06	1.21	0.74
Educational Services	1.73	0.66	0.87	0.69	1.64	0.62	2.02	0.90	0.88	0.92
Health Care and Social Assistance	10.63	0.98	0.99	0.98	1.19	0.83	1.25	0.86	0.95	1.06
Arts, Entertainment, and Recreation	1.24	0.76	0.84	0.72	0.94	1.29	1.17	1.29	1.07	0.94
Accommodation and Food Services	7.36	0.98	0.95	0.96	0.84	1.47	0.94	1.05	1.04	0.94
Other Services	3.90	1.03	1.02	0.97	1.01	0.93	0.89	0.95	1.07	1.00
Public Administration	14.95	1.06	0.91	1.02	0.99	1.06	0.87	1.02	1.06	0.99

Source: Authors' calculations based on agriculture employment data from the BEA's REIS for 1997; Public Administration data from the U.S. Census of Governments, 1997; and the 1998 County Business Patterns.

Table 4: Employment Location Quotients for Canada

	Share of National Employment (percent)	Quebec	Ontario	East	West
Crop and Animal Production	2.68	0.71	0.67	0.77	1.69
Agriculture, Forestry, Fishing and Hunting excluding Crop and Animal Production	0.96	0.92	0.28	4.25	1.20
Mining and Oil and Gas Extraction	1.09	0.41	0.32	0.97	2.29
Utilities	0.76	0.93	1.01	1.05	1.02
Construction	5.64	0.82	0.98	1.12	1.13
Manufacturing	13.96	1.26	1.18	0.78	0.64
Wholesale Trade	4.41	1.04	1.06	0.78	0.95
Retail Trade	11.27	1.01	1.00	1.06	0.98
Transportation and Warehousing	4.97	0.95	0.94	0.98	1.12
Information and Cultural Industries	2.68	0.98	1.07	0.80	0.97
Finance and Insurance	4.08	0.93	1.20	0.68	0.89
Real Estate and Rental and Leasing	1.67	0.81	1.09	0.74	1.09
Professional, Scientific, and Technical Services	6.31	0.94	1.14	0.59	0.97
Management of Companies and Enterprises	0.10	0.83	1.34	0.61	0.80
Administrative and Support and Waste Management and Remediation Services	3.89	0.87	1.10	1.07	0.95
Educational Services	6.55	1.01	0.94	1.05	1.05
Health Care and Social Assistance	9.70	1.05	0.91	1.16	1.03
Arts, Entertainment, and Recreation	1.95	0.90	1.04	0.87	1.05
Accommodation and Food Services	6.72	0.91	0.94	1.02	1.13
Other Services (except Public Administration)	4.80	1.05	0.95	1.09	1.01
Public Administration	5.81	1.08	0.89	1.45	0.96

Source: Authors' calculations based on 2001 Census Bureau data. "East" includes New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland, and Labrador. "West" includes Manitoba, Saskatchewan, Alberta, British Columbia, Yukon Territory, and Northwest Territories.

Table 5: Gini Coefficients for the United States

	1977	1997
Agriculture	0.21	0.13
Mining	0.47	0.50
Construction	0.15	0.10
Manufacturing	0.10	0.10
Transportation	0.06	0.05
Wholesale	0.05	0.05
Retail	0.04	0.03
FIRE	0.07	0.09
Services	0.05	0.03

Source: Authors' calculations using 1977 and 1997 County Business Patterns data for major SIC industries.

Table 6: Summary of U.S. Ellison-Glaeser Indices by Major Sector

	Median	Mean	Minimum	Maximum	Number of Industries
Agriculture, Forestry, Fishing and Hunting excluding Crop and Animal Production	0.058	0.071	0.002	0.207	15
Mining	0.137	0.204	0.002	0.909	29
Utilities	0.013	0.022	0.001	0.091	10
Construction	0.004	0.009	0.001	0.120	28
Manufacturing	0.020	0.041	-0.203	0.585	473
Wholesale Trade	0.004	0.011	-0.005	0.108	69
Retail Trade	0.002	0.005	0.000	0.056	72
Transportation and Warehousing	0.017	0.035	-0.138	0.283	54
Information	0.029	0.052	-0.010	0.314	34
Finance and Insurance	0.009	0.040	-0.041	0.237	38
Real Estate and Rental and Leasing	0.006	0.008	0.000	0.029	24
Professional, Scientific, and Technical Services	0.005	0.016	0.000	0.229	46
Management of Companies and Enterprises	0.003	0.004	0.002	0.009	3
Administrative and Support and Waste Management and Remediation Services	0.004	0.006	-0.002	0.032	43
Educational Services	0.004	0.007	-0.006	0.038	17
Health Care and Social Assistance	0.004	0.006	-0.003	0.034	39
Arts, Entertainment, and Recreation	0.015	0.030	-0.003	0.180	25
Accommodation and Food Services	0.005	0.039	0.000	0.458	15
Other Services	0.003	0.006	-0.001	0.037	48

Source: Authors' calculations based on 1999 County Business Patterns data. Geographic aggregation was at the state level. Indices were computed for every six-digit NAICS industry and then summarized by major sector.

Table 7: Most and Least Concentrated Industries According to EG Index

	Gamma	Geographic Concentration	Plant Herfindahl	Number of Plants
Most Concentrated				
Anthracite mining	0.909	0.912	0.058	64
Sheer hosiery mills	0.585	0.579	0.031	139
Kaolin & ball clay mining	0.561	0.587	0.075	42
Copper ore & nickel ore mining	0.553	0.596	0.121	57
Iron ore mining	0.524	0.569	0.129	36
Wineries	0.498	0.491	0.028	788
Gold ore mining	0.492	0.502	0.060	307
Phosphate rock mining	0.467	0.569	0.196	16
Casino hotels	0.458	0.450	0.012	279
Tobacco stemming & redrying	0.415	0.455	0.085	27
Carpet & rug mills	0.408	0.401	0.172	478
Oil & gas field machinery & equipment	0.394	0.387	0.011	555
Petrochemical mfg.	0.380	0.400	0.048	53
Other hosiery & sock mills	0.333	0.327	0.010	412
Motion picture & video distribution	0.314	0.433	0.201	706
Least Concentrated				
Laminated aluminum foil mfg for flexible pkg	-0.008	0.049	0.058	47
Cyclic crude & intermediate mfg	-0.009	0.053	0.063	57
Scale & balance (except laboratory) mfg	-0.009	0.024	0.034	119
In-vitro diagnostic substance mfg	-0.009	0.089	0.101	223
Radio networks	-0.010	0.113	0.127	339
Explosives mfg	-0.019	0.036	0.055	95
Copper wire (except mechanical) drawing	-0.021	0.041	0.062	67
Cellulose organic fiber mfg	-0.026	0.259	0.279	10
Motor vehicle air-conditioning mfg	-0.026	0.149	0.176	70
House slipper mfg	-0.026	0.178	0.204	20
Cane sugar refining	-0.040	0.073	0.110	19
Monetary authorities - central bank	-0.041	0.020	0.059	46
Other ordnance & accessories mfg	-0.044	0.188	0.230	65
Other urban transit systems	-0.138	0.269	0.365	27
Engineered wood member (exc truss) mfg	-0.203	0.244	0.376	8
Memo:				
Median	0.011			

Source: Authors' calculations based on 1999 County Business Patterns data. Geographic aggregation was at the state level.

Table 8: Summary Statistics for U.S. Manufacturing Belts in 1947 and 1999

1947	United States	Manufacturing Belt	Piedmont	California Region	Rest of the Continental U.S.
Man. Emp	14.304	10.049	0.965	0.602	2.689
Share	1.000	0.702	0.067	0.042	0.188
Est 250+	10,217	7,152	862	378	1,825
Share	1.000	0.700	0.084	0.037	0.178
Total Emp 1950	56.148	27.675	3.098	3.302	22.072
Share	1.000	0.493	0.055	0.059	0.393
LQ man emp	1.000	1.425	1.222	0.715	0.427
LQ est 250	1.000	1.420	1.529	0.629	0.454

1999	United States	Manufacturing Belt	Piedmont	California Region	Rest of the Continental U.S.
Man. Emp	16.559	6.712	1.559	1.571	6.716
Share	1.000	0.405	0.094	0.095	0.406
Est 250+	12,586	4780	1512	945	5349
Share	1.000	0.380	0.120	0.075	0.425
Total Emp 1950	110.088	41.920	7.540	10.620	50.008
Share	1.000	0.381	0.068	0.096	0.454
LQ man emp	1.000	1.064	1.375	0.983	0.892
LQ est 250	1.000	0.997	1.753	0.778	0.936

Note: Employment in millions.

Table 9: Summary Statistics for City Size Groups

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Maximum Neighboring Population (millions)	0.421	1.369	3.426	15.369
Number of Counties	2,432	468	151	59
Percent of Land Area	87.2	8.3	3.6	0.9
Percent of Employment	25.0	25.0	24.8	25.2
Percent of Population	27.7	25.3	23.6	23.5
Employment as a percent of population	53.0	58.0	61.8	62.9
Percent of population 65 or older	14.0	12.6	11.6	11.2
Percent college graduate of population 25-34 years old	18.5	25.6	31.0	34.6

Note: The data on land area, population, and population characteristics are all from the 2000 decennial census.

Table 10: Distribution of Employment Across Density Quartiles
Major Sectors Sorted By Urbanness of Industry

NAICS	Description	Share of Total Employment	LQ1	LQ2	LQ3	LQ4	Sector Urbanization (LQ4-LQ1)
Rural							
111-112	Crop and Animal Production	0.019	2.75	0.85	0.31	0.08	-2.67
113-115	Agriculture, Forestry, Fishing and Hunting excluding Crop and Animal Production	0.001	2.60	0.91	0.34	0.16	-2.43
21	Mining	0.002	2.50	0.78	0.28	0.44	-2.06
31	Manufacturing	0.100	1.18	1.13	0.93	0.76	-0.43
22	Utilities	0.004	1.22	1.04	0.90	0.84	-0.39
-	Government	0.138	1.22	1.04	0.89	0.86	-0.36
Diffuse							
44	Retail trade	0.090	1.08	1.07	1.00	0.86	-0.22
-	Proprietorships	0.156	1.13	0.97	0.97	0.93	-0.21
72	Accommodation & food services	0.060	1.03	1.04	1.06	0.87	-0.16
62	Health care and social assistance	0.085	1.04	1.01	0.96	0.99	-0.05
23	Construction	0.040	0.89	1.08	1.14	0.89	0.00
99	Unclassified establishments	0.001	1.07	0.85	0.96	1.11	0.04
71	Arts, entertainment & recreation	0.011	0.84	0.97	1.17	1.03	0.19
81	Other services (except public administration)	0.032	0.87	1.05	1.02	1.07	0.20
Urban							
48	Transportation & warehousing	0.023	0.73	0.93	1.03	1.31	0.57
42	Wholesale trade	0.037	0.66	0.93	1.14	1.26	0.61
53	Real estate & rental & leasing	0.012	0.67	0.90	1.12	1.31	0.65
95	Auxiliaries (exc corporate, subsidiary & regional mgt)	0.006	0.61	0.89	1.20	1.30	0.69
56	Admin, support, waste mgt, remediation services	0.053	0.57	1.00	1.18	1.26	0.69
52	Finance & insurance	0.036	0.62	0.92	1.13	1.33	0.71
61	Educational services	0.015	0.68	0.94	0.98	1.40	0.72
51	Information	0.021	0.57	0.81	1.14	1.48	0.91
55	Management of companies & enterprises	0.017	0.35	0.90	1.27	1.48	1.13
54	Professional, scientific & technical services	0.041	0.46	0.78	1.11	1.64	1.17

Table 11: Distribution of Employment Across Density Quartiles
Top 10 and Bottom 10 Detailed Industries Sorted By Urbanness of Industry

NAICS	Description	LQ1	LQ2	LQ3	LQ4	Industry Urbanization (LQ4-LQ1)
Bottom Ten						
212210	Iron ore mining	3.94	0.03	0.03	0.00	-3.93
212111	Bituminous coal & lignite surface mining	3.71	0.20	0.09	0.01	-3.70
212222	Silver ore mining	3.67	0.31	0.01	0.01	-3.66
321212	Softwood veneer & plywood mfg	3.65	0.28	0.08	0.00	-3.65
212324	Kaolin & ball clay mining	3.49	0.41	0.10	0.00	-3.49
325182	Carbon black mfg	3.60	0.17	0.06	0.17	-3.43
115111	Cotton ginning	3.43	0.47	0.09	0.01	-3.42
315291	Infants' cut & sew apparel mfg	3.47	0.40	0.05	0.08	-3.39
212112	Bituminous coal underground mining	3.36	0.49	0.14	0.01	-3.35
113310	Logging	3.37	0.48	0.13	0.02	-3.35
Top Ten						
512199	Other motion picture & video industries	0.20	0.32	0.45	3.02	2.82
512191	Teleproduction & oth postproduction services	0.09	0.29	0.66	2.95	2.86
523110	Investment banking & securities dealing	0.05	0.24	0.73	2.96	2.91
523130	Commodity contracts dealing	0.11	0.31	0.41	3.15	3.04
522293	International trade financing	0.09	0.10	0.54	3.25	3.16
512120	Motion picture & video distribution	0.07	0.30	0.29	3.31	3.24
483112	Deep sea passenger transportation	0.03	0.30	0.19	3.44	3.41
512220	Integrated record production, distribution	0.03	0.25	0.21	3.48	3.45
485112	Commuter rail systems	0.17	0.00	0.04	3.75	3.57
523210	Securities & commodity exchanges	0.00	0.00	0.06	3.90	3.90

Table 12: Distribution of Employment By Groups and Urbanness Quartile

NAICS	Share of Total Employment	LQ1	LQ2	LQ3	LQ4	Industry Urbanization (LQ4-LQ1)
Major Sectors						
Rural	0.264	1.33	1.06	0.86	0.76	-0.57
Diffuse	0.474	1.05	1.02	1.01	0.93	-0.12
Urban	0.262	0.58	0.91	1.13	1.38	0.79
Narrow Industries						
Rural	0.282	1.45	1.06	0.81	0.68	-0.78
Diffuse	0.451	1.02	1.04	1.02	0.93	-0.09
Urban	0.266	0.50	0.88	1.16	1.46	0.96

Table 13: Location Quotients for Canada

NAICS	Description	Share of Total Employment	LQ1	LQ2	LQ3	LQ4	Sector Urbanization (LQ4-LQ1)
113-115	Agriculture, Forestry, Fishing and Hunting excluding Crop and Animal Production	0.009	3.24	0.41	0.38	0.05	-3.19
21	Mining and Oil and Gas Extraction	0.012	2.07	0.64	1.38	0.09	-1.99
111-112	Crop and Animal Production	0.011	2.07	1.25	0.63	0.13	-1.94
61	Educational Services	0.046	1.77	1.02	0.76	0.50	-1.27
91	Public Administration	0.039	1.06	1.50	0.97	0.52	-0.54
62	Health Care and Social Assistance	0.093	1.26	1.09	0.93	0.75	-0.51
48-49	Transportation and Warehousing	0.039	1.13	1.06	1.07	0.78	-0.34
72	Accommodation and Food Services	0.084	1.14	1.03	1.04	0.82	-0.32
44-45	Retail Trade	0.117	1.10	1.02	1.03	0.87	-0.22
23	Construction	0.064	1.05	1.05	1.08	0.85	-0.21
71	Arts, Entertainment and Recreation	0.018	0.96	0.89	1.25	0.93	-0.03
81	Other Services (except Public Administration)	0.041	0.97	1.05	1.04	0.95	-0.02
22	Utilities	0.005	0.73	1.53	0.83	0.89	0.16
31-33	Manufacturing	0.145	0.90	1.07	0.85	1.14	0.24
53	Real Estate and Rental and Leasing	0.025	0.74	0.90	1.24	1.14	0.40
55	Management of Companies and Enterprises	0.016	0.60	0.89	1.11	1.36	0.75
52	Finance and Insurance	0.044	0.60	1.01	0.89	1.43	0.84
56	Administrative and Support, Waste Management and Remediation Services	0.043	0.59	0.83	1.05	1.47	0.88
41	Wholesale Trade	0.061	0.62	0.84	0.97	1.51	0.89
54	Professional, Scientific and Technical Services	0.065	0.40	0.73	1.34	1.50	1.10
51	Information and Cultural Industries	0.022	0.49	0.71	1.06	1.66	1.17

Table 14: Variation in Industry Characteristics by Degree of Concentration

Panel A. Plant Size (Employment per plant)

	All Industries		Manufacturing Industries Only	
Quartile	Average Size	Mean EG	Average Size	Mean Gamma
1	6.6	0.015	25.0	0.027
2	16.1	0.025	48.4	0.035
3	41.2	0.034	80.3	0.040
4	167.2	0.060	268.0	0.062

Panel B. Materials Intensity (Materials as a fraction of total sales)

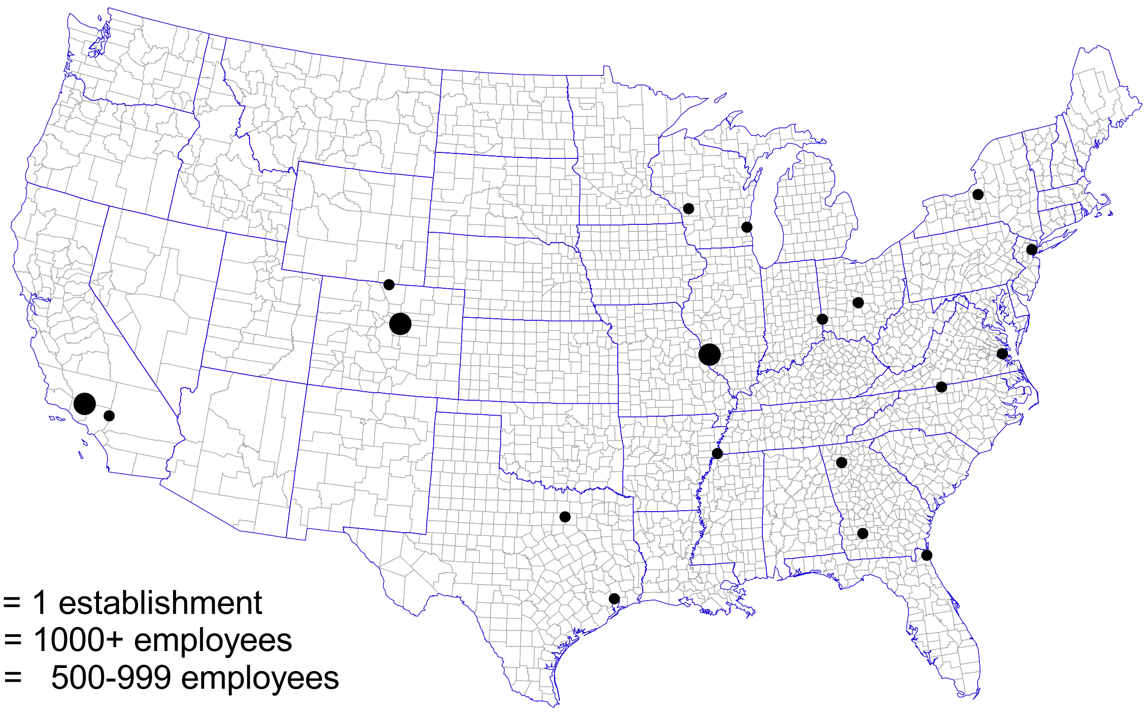
	Manufacturing Industries	
Quartile	Mean	Mean EG
1	0.35	0.027
2	0.45	0.032
3	0.52	0.040
4	0.66	0.065

Note: Panel B can be constructed only for manufacturing. "EG" refers to the Ellison-Glaeser (1997) index of geographic concentration.

Table 15: Location and Market Areas of Anheuser-Busch Plants

Plant Location	Average distance of market	Estimated Share of Market	Neighboring Population percentile of location
Fairfield, Calif.	285.8	0.084	0.551
Van Nuys, Calif.	125.8	0.099	0.952
Fort Collins, Colo.	319.4	0.050	0.282
Houston, Tex.	207.7	0.097	0.788
St. Louis, Mo.	270.7	0.143	0.667
Cartersville, Ga.	153.2	0.084	0.479
Columbus, Ohio	163.5	0.130	0.578
Jacksonville, Fla.	197.7	0.064	0.450
Williamsburg, Va.	135.2	0.067	0.366
Baldwinsville, N.Y.	93.3	0.021	0.391
Newark, N.J.	50.1	0.115	0.979
Merrimack, N.H.	70.6	0.046	0.408

Figure 1:
Location of Large Breweries (CBP 1999)



1 dot = 1 establishment
● = 1000+ employees
• = 500-999 employees

Figure 2:
Location of Large Manufacturing Plants (1947)

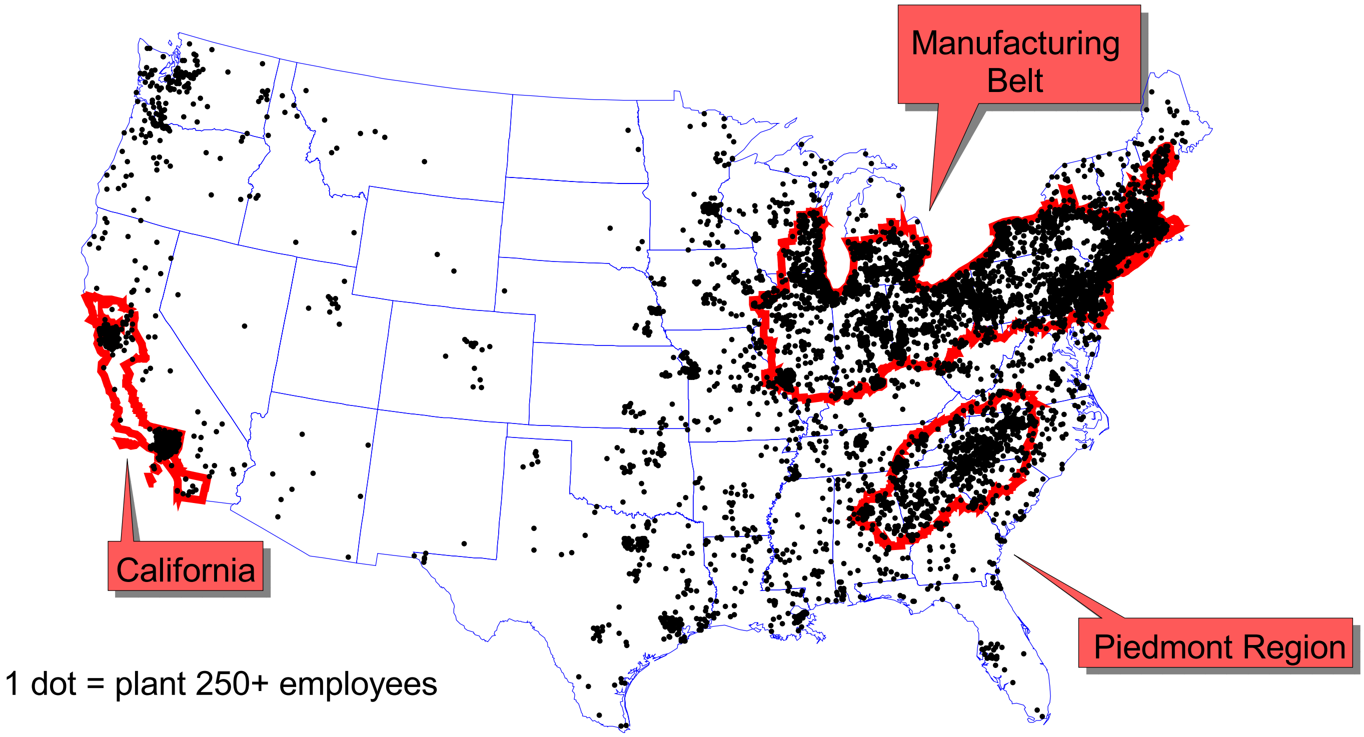


Figure 3:
Location of Large Manufacturing Plants (1999)

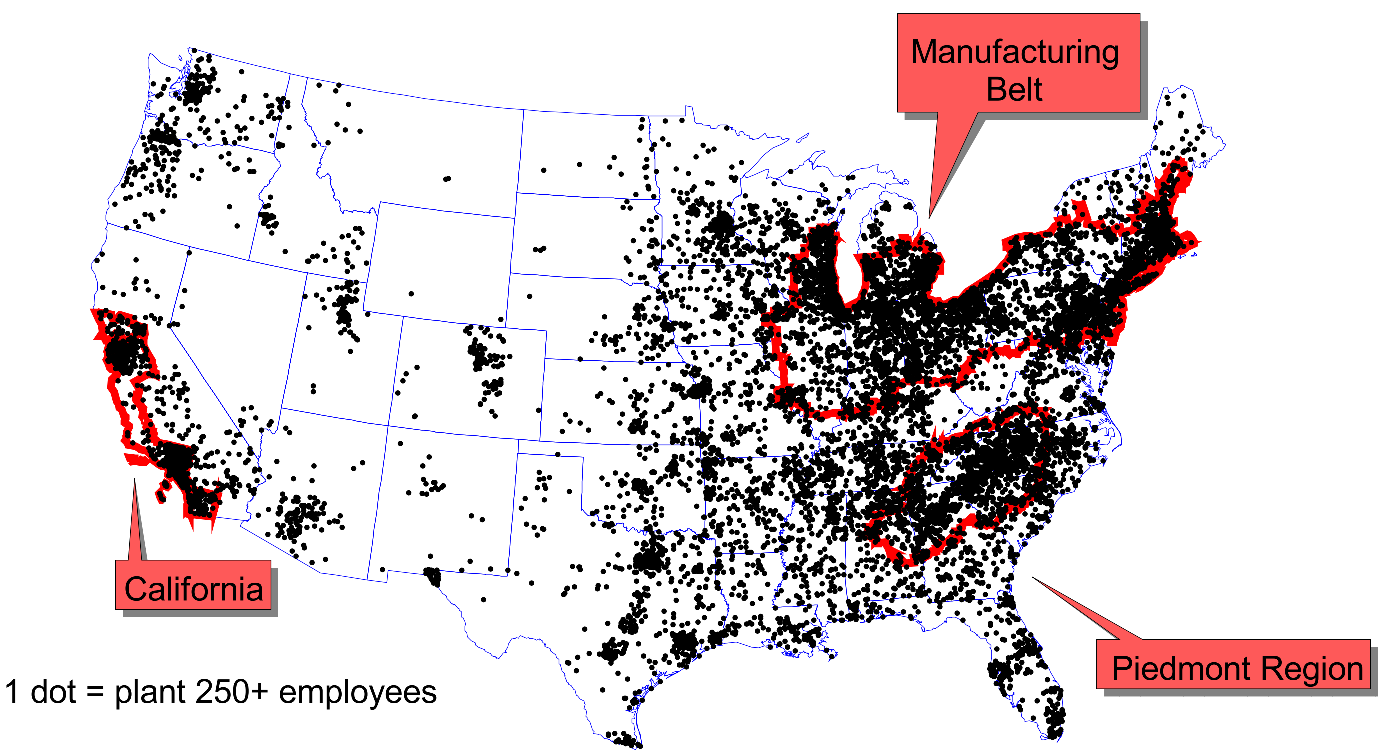
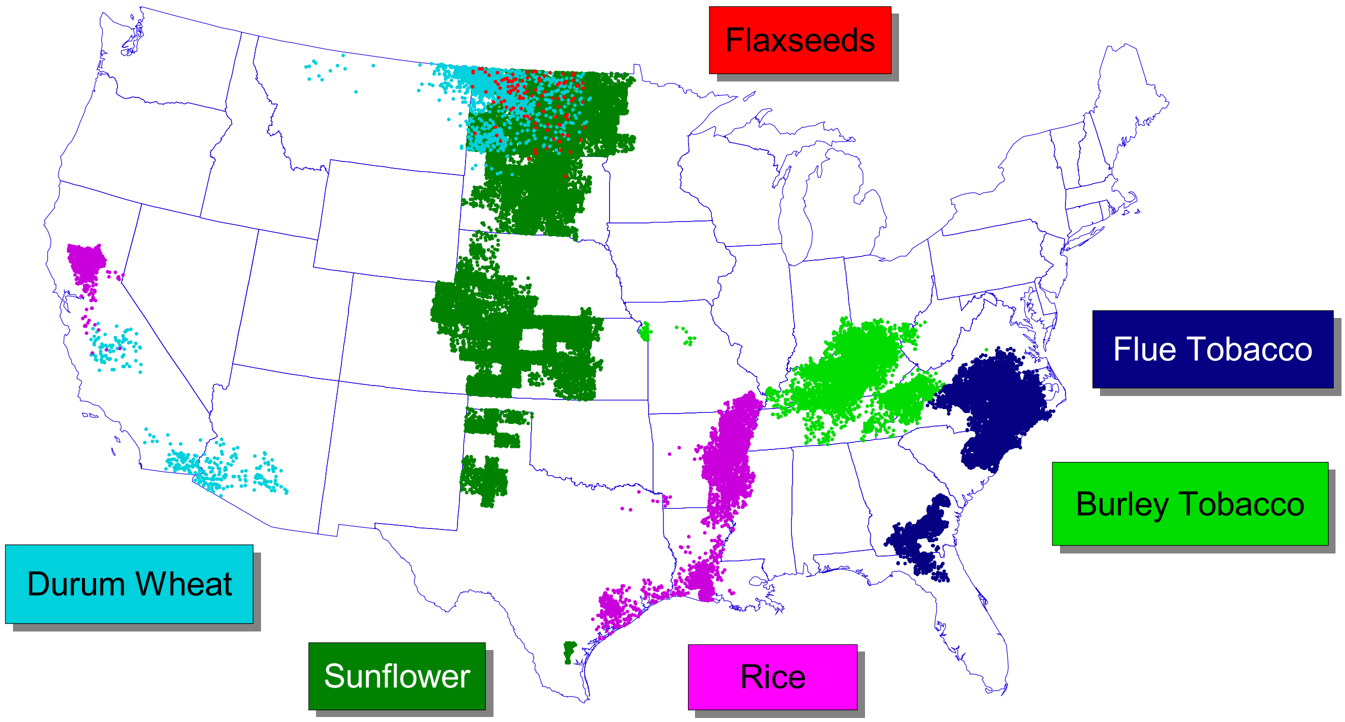
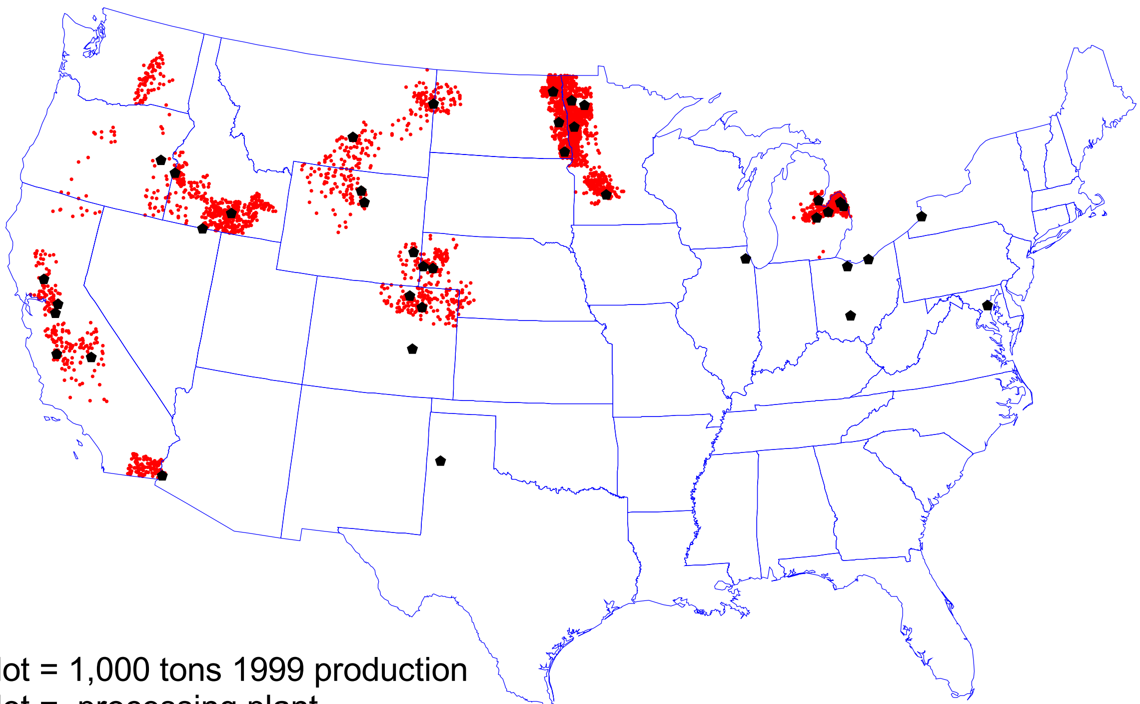


Figure 4:
Location of Durum Wheat, Rice, Flue Tobacco, and Burley Tobacco



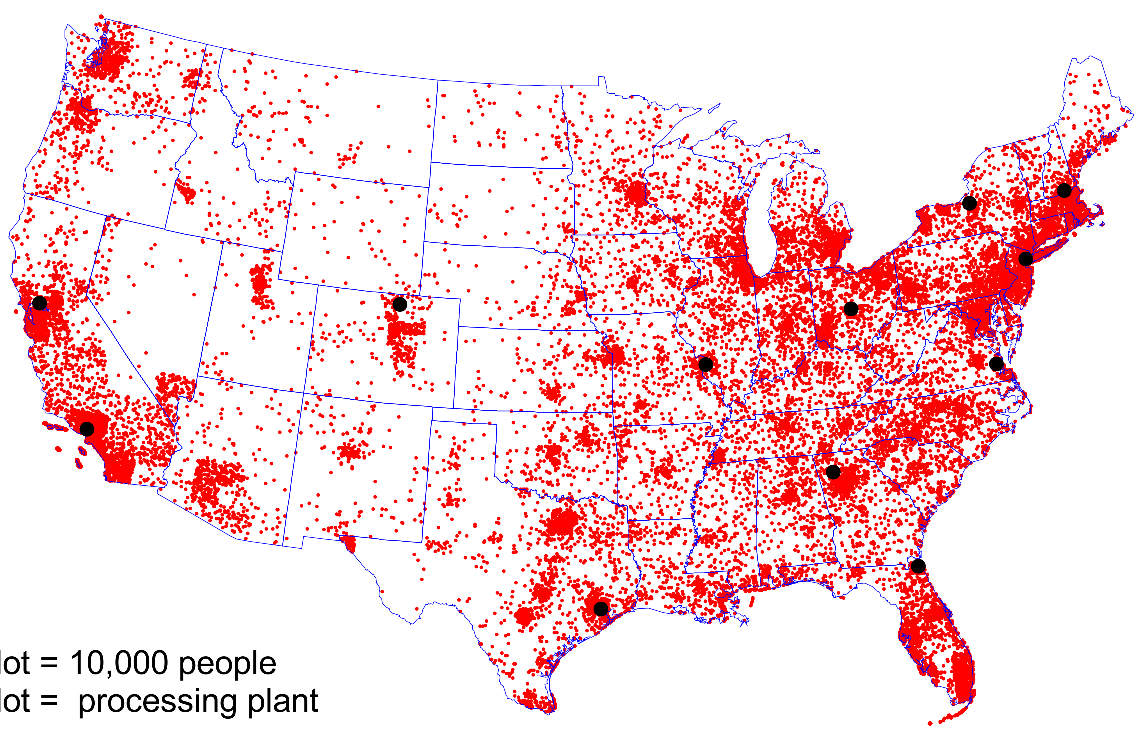
1 dot = 50,000 tons

Figure 5:
Location of Sugar Beet Plants and Sugar Beet Crops



● 1 dot = 1,000 tons 1999 production
● 1 dot = processing plant

Figure 6:
Location of Anheuser-Busch Breweries and Population (2000)



● 1 dot = 10,000 people
● 1 dot = processing plant

Figure 7:
Location of Consumer Lending (NAICS 522291)
(Location Quotient by County)