CHAPTER 1

Introduction and Axioms of Urban Economics

Cities have always been the fireplaces of civilization, whence light and heat radiated out into the dark.

—THEODORE PARKER

I'd rather wake up in the middle of nowhere than in any city on earth.

—STEVE McQUEEN

This book explores the economics of cities and urban problems. The quotes from Parker and McQueen reflect our mixed feelings about cities. On the positive side, cities facilitate innovation; production, and trade; so they increase our standard of living. On the negative side, cities are noisy, dirty, and crowded. As we'll see in the first part of the book, firms and people locate in cities because the obvious costs of being in a city are more than offset by subtle benefits of producing in close proximity to other firms and people. As we'll see later in the book, policies that combat urban problems such as congestion, pollution, and crime are likely to increase the vitality of cities, causing them to grow.

WHAT IS URBAN ECONOMICS?

The discipline of urban economics is defined by the intersection of geography and economics. Economics explores the choices people make when resources are limited. Households make choices to maximize their utility, while firms maximize their profit. Geographers study how things are arranged across space, answering the question, Where does human activity occur? Urban economics puts economics and geography together, exploring the geographical or location choices of utility-maximizing households and profit-maximizing firms. Urban economics also identifies inefficiencies in location choices and examines alternative public policies to promote efficient choices.
Urban economics can be divided into six related areas that correspond to the six parts of this book.

1. **Market forces in the development of cities.** The interurban location decisions of firms and households generate cities of different size and economic structure. We explore the issues of why cities exist and why there are big cities and small ones.

2. **Land use within cities.** The intrasurban location decisions of firms and households generate urban land-use patterns. In modern cities, employment is spread throughout the metropolitan area, in sharp contrast to the highly centralized cities of just 100 years ago. We explore the economic forces behind the change from centralized to decentralized cities. We also use a model of neighborhood choice to explore the issue of segregation with respect to race, income, and educational level.

3. **Urban transportation.** We explore some possible solutions to the urban congestion problem and look at the role of mass transit in the urban transportation system. One issue is whether a bus system is more efficient than a light-rail system or a heavy-rail system like BART (San Francisco) or Metro (Washington).

4. **Crime and public policy.** We look at the problem of urban crime and show the links between crime and two other urban problems, poverty and low educational achievement.

5. **Housing and public policy.** Housing choices are linked to location choices because housing is immobile. We’ll discuss why housing is different from other products and how housing policies work.

6. **Local government expenditures and taxes.** Under our fragmented system of local government, most large metropolitan areas have dozens of local governments, including municipalities, school districts, and special districts. In making location choices, households consider the mix of taxes and local public goods.

**WHAT IS A CITY?**

An urban economist defines an urban area as a geographical area that contains a large number of people in a relatively small area. In other words, an urban area has a population density that is high relative to the density of the surrounding area. This definition accommodates urban areas of vastly different sizes, from a small town to a large metropolitan area. The definition is based on population density because an essential feature of an urban economy is frequent contact between different economic activities which is feasible only if firms and households are concentrated in a relatively small area.

The U.S. Census Bureau has developed a variety of geographical definitions relevant to urban economics. Since much of the empirical work in urban economics is based on census data, a clear understanding of these definitions is important. The appendix to this chapter provides the details of the census definitions.

The key census definitions, some of which are new for the 2000 Census, are as follows.

1. **Urban area:** A densely settled geographical area with a minimum population of 2,500 people and a minimum density of 5,000 people per square mile. In 2000, there were 3,756 urban areas in the United States.

2. **Urban population:** People living in urban areas. In 2000, the urban population was 79 percent of the total population.

3. **Metropolitan area:** A core area with a substantial population nucleus, together with adjacent communities that are integrated, in an economic sense, with the core area. To qualify as a metropolitan area, the minimum population is 50,000 people. In 2000, there were 361 metropolitan statistical areas in the United States.

4. **Micropolitan area:** A smaller version of a metropolitan area with a concentration of 10,000 to 50,000 people. In 2000, there were 559 micropolitan statistical areas in the United States.

5. **Principal city:** The largest municipality in each metropolitan or micropolitan statistical area. A municipality is defined as an area over which a municipal corporation exercises political authority and provides local government services such as sewage service, crime protection, and fire protection.

This book uses three terms to refer to spatial concentrations of economic activity: urban areas, metropolitan areas, and cities. These three terms, which will be used interchangeably, refer to the economic activity (an area with a relatively high population density that contains a set of closely related activities), not the political city. When referring to a political city, we will use the term central city or municipality.

**WHY DO CITIES EXIST?**

This is the fundamental question of urban economics. People need land to produce food and other resources, and living in dense cities separates us from the land where food is produced. As Burtlett (1998) points out, no other creatures in the animal world form anything like cities. Herbivores such as wildebeests and bison form larger herds but constantly migrate to fresh land to ensure a steady supply of food. Coral is concentrated in stationary reefs, but ocean currents provide a steady supply of food to the stationary coral. Perhaps the closest thing to a city in the natural world is a bee hive or an ant hill. Eusocial insects such as bees and ants form colonies with thousands of inhabitants, with highly specialized castes—soldier ants, drones, breeders, nurses, and cleanup crew. In contrast with human cities, these insect aggregations are closed to non-natives and not based on voluntary exchange.

Cities exist because human technology has created systems of production and exchange that seem to defy the natural order. Three conditions must be satisfied for a city to develop.

1. **Agricultural surplus.** People outside cities must produce enough food to feed themselves and city dwellers.
2. **Urban production.** City dwellers must produce something—goods or services—to exchange for food grown by rural workers.

3. **Transportation for exchange.** There must be an efficient transportation system to facilitate the exchange of food and urban products.

Figure 1–1 shows the share of people living in cities in the United States from 1800 to 2000. Over this period, the urban share increased from 6 percent to 79 percent, a remarkable transformation that also occurred in other parts of the world. As we'll see in the next three chapters of the book, the transformation of a rural society into an urban one occurred because technological advances increased the agricultural surplus (condition 1), increased the productivity of urban workers (condition 2), and increased the efficiency of transportation and exchange (condition 3).

Figure 1–2 shows urbanization rates for different regions around the world, with projections for the year 2030. In 1950, urbanization rates were relatively low in Africa and Asia, and highest in Oceania and North America. Between now and the year 2030, urbanization rates are expected to increase everywhere, with the largest increases in Africa and Asia. For the world as a whole, the urbanization rate was 30 percent in 1950 and is expected to double by the year 2030.

Table 1–1 (page 6) shows the population figures for the nation’s 30 largest metropolitan areas. The New York area tops the list, followed by Los Angeles, Chicago, Philadelphia, and Dallas. The third column shows the percentage growth of each metropolitan area over the period 1990 to 2000. The most rapidly growing metropolitan areas were in the South, the Mountain States, and the West. Over the longer period 1980–2000, two other metropolitan areas, Cleveland and Baltimore, lost population. These metropolitan areas experienced large losses in manufacturing employment, including losses in primary metals, motor vehicles, rubber, and non-electrical machinery industries.

Table 1–2 (page 7) shows the population figures for the world’s largest metropolitan areas outside the United States. The table shows actual populations in 1995 and projected populations for the year 2015. Six metropolitan areas, all of which are in the developing world, are expected to grow by at least 50 percent over the 20-year period. In contrast, three cities in the developed world (Tokyo, Osaka, and Paris) are expected to grow at moderate to slow rates. In the United States, New York is expected to grow 8 percent over the period, and Los Angeles is expected to grow 15 percent.
TABLE 1–1  Largest Metropolitan Areas in the United States, 2000

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New York-Northern New Jersey-Long Island, NY-NJ-PA</td>
<td>18,323,002</td>
<td>8.8</td>
</tr>
<tr>
<td>Los Angeles-Long Beach-Santa Ana, CA</td>
<td>12,362,077</td>
<td>9.7</td>
</tr>
<tr>
<td>Chicago-Naperville-Joliet, IL-IN-WI</td>
<td>9,088,316</td>
<td>11.2</td>
</tr>
<tr>
<td>Philadelphia-Camden-Wilmington, PA-NJ-DE</td>
<td>5,687,147</td>
<td>4.6</td>
</tr>
<tr>
<td>Dallas-Fort Worth-Arlington, TX</td>
<td>5,161,534</td>
<td>29.4</td>
</tr>
<tr>
<td>Miami-Fort Lauderdale-Miami Beach, FL</td>
<td>5,007,566</td>
<td>23.5</td>
</tr>
<tr>
<td>Washington-Arlington-Alexandria, DC-VA-MD</td>
<td>4,796,183</td>
<td>16.3</td>
</tr>
<tr>
<td>Houston-Baytown-Sugar Land, TX</td>
<td>4,715,407</td>
<td>25.2</td>
</tr>
<tr>
<td>Detroit-Warren-Livonia, MI</td>
<td>4,452,257</td>
<td>4.8</td>
</tr>
<tr>
<td>Boston-Cambridge-Madison, MA-NH</td>
<td>4,391,344</td>
<td>6.2</td>
</tr>
<tr>
<td>Atlanta-Sandy Springs-Marietta, GA</td>
<td>4,247,981</td>
<td>38.4</td>
</tr>
<tr>
<td>San Francisco-Oakland- Fremont, CA</td>
<td>4,123,740</td>
<td>11.9</td>
</tr>
<tr>
<td>Riverside-San Bernardino- Ontario, CA</td>
<td>3,254,821</td>
<td>22.7</td>
</tr>
<tr>
<td>Phoenix-Mesa-Scottsdale, AZ</td>
<td>3,201,876</td>
<td>45.3</td>
</tr>
<tr>
<td>Seattle-Tacoma-Bellevue, WA</td>
<td>3,043,878</td>
<td>18.9</td>
</tr>
<tr>
<td>Minneapolis-St Paul-Bloomington, MN-WI</td>
<td>2,968,806</td>
<td>16.9</td>
</tr>
<tr>
<td>San Diego-Carlsbad-San Marcos, CA</td>
<td>2,813,833</td>
<td>12.6</td>
</tr>
<tr>
<td>St Louis, MO-IL</td>
<td>2,698,687</td>
<td>4.6</td>
</tr>
<tr>
<td>Baltimore-Towson, MD</td>
<td>2,592,994</td>
<td>7.2</td>
</tr>
<tr>
<td>Pittsburgh, PA</td>
<td>2,431,067</td>
<td>1.1</td>
</tr>
<tr>
<td>Tampa-St Petersburg-Clearwater, FL</td>
<td>2,395,997</td>
<td>15.9</td>
</tr>
<tr>
<td>Denver-Aurora, CO</td>
<td>2,179,240</td>
<td>30.7</td>
</tr>
<tr>
<td>Cleveland-Elyria-Mentor, OH</td>
<td>2,148,143</td>
<td>2.2</td>
</tr>
<tr>
<td>Cincinnati-Middletown, OH-KY-IN</td>
<td>2,009,632</td>
<td>8.9</td>
</tr>
<tr>
<td>Portland-Vancouver-Beaverton, OR-WA</td>
<td>1,927,881</td>
<td>26.5</td>
</tr>
<tr>
<td>Kansas City, MO-KS</td>
<td>1,816,038</td>
<td>12.2</td>
</tr>
<tr>
<td>Sacramento-Aden-Arco-Roseville, CA</td>
<td>1,794,857</td>
<td>21.3</td>
</tr>
<tr>
<td>San Jose-Sunnyvale-Santa Clara, CA</td>
<td>1,735,819</td>
<td>13.1</td>
</tr>
<tr>
<td>San Antonio, TX</td>
<td>1,711,703</td>
<td>21.6</td>
</tr>
<tr>
<td>Orlando, FL</td>
<td>1,644,361</td>
<td>34.3</td>
</tr>
</tbody>
</table>


Figure 1–3 (page 8) shows the time trend of large urban agglomerations in the world, defined as metropolitan areas with at least 1 million people. The figure distinguishes between cities in the developed and less developed regions. In 1970, the two types of regions had roughly the same number of large cities. By 1996, however, the number of large cities in the less developed regions nearly doubled, and by 2015 there will be roughly four times as many large cities in less developed regions.

THE FIVE AXIOMS OF URBAN ECONOMICS

Urban economics explores the location choices of households and firms, and so it is natural to assume that people and firms are mobile. Of course, people don’t instantly change their workplaces and residences when circumstances change; therefore, a model of perfect mobility tells us more about long-term changes than short-term ones. The average household changes its residence every seven years, meaning that about 14 percent of the population moves every year. Although most models of urban economics assume perfect mobility, there are exceptions, and we will highlight the analysis that assumes less than perfect mobility.

In this part of the chapter, we introduce five axioms of urban economics. An axiom is a self-evident truth, something that most people readily understand and accept. For our purposes, “most people” are people who have taken at least one course in economics. The five axioms lie at the heart of urban economics and together provide a foundation for the economic models of location choices. As you go through the book, these five axioms will appear repeatedly.

1. Prices Adjust to Achieve Location Equilibrium

A location equilibrium occurs when no one has an incentive to move. Suppose that you and Bud are competing for two rental houses, one along a beautiful beach and one along a noisy highway. If the two houses have the same price (the same monthly rent), you would prefer the beach house, and so would Bud. Flipping a coin and giving the beach house to the winner wouldn’t generate a location equilibrium because the unlucky person in the highway house would have an incentive to move to the more desirable house.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo, Japan</td>
<td>26,96</td>
<td>28,89</td>
<td>7.6</td>
</tr>
<tr>
<td>Mexico City, Mexico</td>
<td>16,56</td>
<td>19,18</td>
<td>16</td>
</tr>
<tr>
<td>Sao Paulo, Brazil</td>
<td>16,53</td>
<td>20,32</td>
<td>23</td>
</tr>
<tr>
<td>Bombay, India</td>
<td>15,14</td>
<td>26,22</td>
<td>73</td>
</tr>
<tr>
<td>Shanghai, China</td>
<td>13,58</td>
<td>17,97</td>
<td>32</td>
</tr>
<tr>
<td>Calcutta, India</td>
<td>11,92</td>
<td>17,31</td>
<td>45</td>
</tr>
<tr>
<td>Buenos Aires, Argentina</td>
<td>11,80</td>
<td>13,86</td>
<td>17</td>
</tr>
<tr>
<td>Seoul, Korea</td>
<td>11,61</td>
<td>12,98</td>
<td>12</td>
</tr>
<tr>
<td>Beijing, China</td>
<td>11,30</td>
<td>13,57</td>
<td>38</td>
</tr>
<tr>
<td>Osaka, Japan</td>
<td>10,61</td>
<td>10,61</td>
<td>0</td>
</tr>
<tr>
<td>Lagos, Nigeria</td>
<td>10,29</td>
<td>24,61</td>
<td>139</td>
</tr>
<tr>
<td>Rio de Janeiro, Brazil</td>
<td>10,18</td>
<td>11,86</td>
<td>17</td>
</tr>
<tr>
<td>Delhi, India</td>
<td>9,95</td>
<td>16,86</td>
<td>69</td>
</tr>
<tr>
<td>Karachi, Pakistan</td>
<td>9,73</td>
<td>19,38</td>
<td>99</td>
</tr>
<tr>
<td>Cairo, Egypt</td>
<td>9,49</td>
<td>14,42</td>
<td>49</td>
</tr>
<tr>
<td>Paris, France</td>
<td>9,32</td>
<td>9,69</td>
<td>2</td>
</tr>
<tr>
<td>Tianjin, China</td>
<td>9,42</td>
<td>13,53</td>
<td>44</td>
</tr>
<tr>
<td>Manila, Philippines</td>
<td>9,39</td>
<td>14,65</td>
<td>58</td>
</tr>
<tr>
<td>Moscow, Russia Fed.</td>
<td>9,27</td>
<td>9,30</td>
<td>0</td>
</tr>
<tr>
<td>Jakarta, Indonesia</td>
<td>8,62</td>
<td>13,92</td>
<td>61</td>
</tr>
<tr>
<td>Dhaka, Bangladesh</td>
<td>8,55</td>
<td>19,49</td>
<td>128</td>
</tr>
</tbody>
</table>

The price of land also adjusts to ensure locational equilibrium among firms. Office firms compete for the most accessible land in a city, and land at the center is the most accessible and thus the most expensive. In equilibrium, office firms on less accessible land far from the center pay lower prices for land, and can be just as profitable as firms on the most accessible land.

2. Self-Reinforcing Effects Generate Extreme Outcomes.

A self-reinforcing effect is a change in something that leads to additional changes in the same direction. Consider a city where the sellers of new automobiles are initially spread evenly throughout the city. If one seller relocates next to another seller on Auto Road, what happens next? Auto consumers compare brands before buying, and the pair of sellers on Auto Road will facilitate comparison shopping and thus attract buyers. The increased consumer traffic on Auto Road will make it an attractive site for other auto sellers, so they will move too. The ultimate result is an "auto row," a cluster of firms that compete against one another, yet locate nearby.

Self-reinforcing changes also happen in the location decisions of people. Suppose artists and creative types are initially spread out evenly across a dozen cities in a region. If by chance one city experiences an influx of artists, its creative environment will improve as artists (1) are exposed to more ideas and fabrication techniques and (2) can share studios, print shops, tool suppliers, and other facilities. The cluster of artists will attract other artists from the region, causing a concentration of artistic production in one city. In recent decades, cities that have attracted artists and creative folks have experienced relatively rapid growth (Florida, 2002).

3. Externalities Cause Inefficiency

In most transactions, the costs and benefits of the exchange are confined to the individual buyer and seller. The consumer pays a price equal to the full cost of producing the good, so no one else bears a cost from the transaction. Similarly, the consumer is the only person to benefit from the product. In contrast, an externality occurs when some of the costs or benefits of a transaction are experienced by someone other than the buyer or seller, that is, someone external to the transaction.

An external cost occurs when a consumer pays a price that is less than the full cost of producing a product. The price of a product always includes the costs of the labor, capital, and raw materials used to produce the product, but it usually does not include the environmental costs of producing the product. For example, if burning gasoline in automobiles generates air pollution, part of the cost of driving is borne by people who breathe dirty air. Similarly, when you enter a crowded highway, you slow down everyone else, meaning that other drivers bear a cost.

An external benefit occurs when a product purchased by one person generates a benefit for someone else. For example, painting my peeling house improves the appearance of my neighborhood, increasing the value of my neighbor’s house as well as mine. Education generates external benefits because it improves communication and thinking skills, making a person a better team worker. In other
words, some of the benefits of education are experienced by a person's fellow workers, who become more productive and thus earn higher wages.

When there are external costs or benefits, we do not expect the market equilibrium to be socially efficient. In the case of external cost, people pay less than the full social cost of an action like driving, so they drive too much. In the case of external benefit, people get less than the full social benefit from an action like education, so they stop short of the socially efficient level of education. As we'll see later in the book, cities have all sorts of external costs and benefits. In many cases there is a simple solution: Internalize the externality with a tax or a subsidy, and let individuals, who then bear the full social cost and benefits of their actions, decide what to do.

4. Production Is Subject to Economies of Scale

Economies of scale occur when the average cost of production decreases as output increases. For most products, if we start with a relatively small production operation and double all inputs, the average cost of production decreases. In the jargon of economics, when the long-run average cost curve is negatively sloped, we say that there are scale economies in production. Scale economies occur for two reasons:

- **Indivisibilities.** Some capital inputs are "lumpy" and cannot be scaled down for small operations. As a result, a small operation has the same indivisible inputs as a large operation. For example, to manufacture frisbees you need a mold, whether you produce one frisbee per day or a thousand. Similarly, to produce microprocessors you need a clean room and other expensive equipment, whether you produce one processor per day or a thousand. As output increases, the average cost decreases because the cost of the indivisible input is spread over more output.

- **Factor specialization.** In a small one-person production operation, a worker performs a wide variety of production tasks. In a larger operation with more workers, each worker specializes in a few tasks, leading to higher productivity because of continuity (less time is spent switching from one task to another) and proficiency (from experience and learning). The notion of factor specialization is captured in the old expression, "A jack of all trades is master of none." Adding to this expression, we can say that a specialized worker is a master of one task.

As we'll see later in the book, scale economies play a vital role in urban economics. In fact, as we'll see in Chapter 2, if there are no scale economies, there will be no cities. It is costly to transport products from a production site to consumers, so centralized production in cities will be sensible only if there is some advantage that more than offsets transport costs.

The extent of scale economies in production varies across products. Microprocessors are produced in $5 billion fabrication facilities with a highly specialized workforce performing hundreds of complex tasks, resulting in large scale economies in production. In contrast, pizza is produced with a $5,000 pizza oven with just a few production tasks, so scale economies are exhausted sooner. In general, the extent of scale economies is determined by the lumpiness of indivisible inputs and the opportunities for factor specialization.

5. Competition Generates Zero Economic Profit

When there are no restrictions on the entry of firms into a market, we expect firms to enter the market until economic profit is zero. Recall that economic profit equals the excess of total revenue over total economic cost, where economic cost includes the opportunity costs of all inputs. Two key components of economic costs are the opportunity cost of the entrepreneur's time and the opportunity cost of funds invested in the firm. For example, suppose an entrepreneur could earn $50,000 in another job and invests $100,000 in the firm, taking the money out of a mutual fund that earns 8 percent. The economic cost of the firm includes $60,000 in time cost and $8,000 in investment cost. Once we account for all the opportunity costs, the fact that economic profit is zero means that a firm is making enough money to stay in business, but not enough for other firms to enter the market. Earning zero economic profit means earning "normal" accounting profit.

In urban economics, competition has a spatial dimension. Each firm enters the market at some location, and the profit of each firm is affected by the locations of other firms. Spatial competition looks a lot like monopolistic competition, a market structure in which firms sell slightly differentiated products in an environment of unrestricted entry. Although this sounds like an oxymoron such as "tight slacks" and "jumbo shrimp," the words are revealing. Each firm has a monopoly for its differentiated product, but unrestricted entry leads to keen competition for consumers who can easily switch from one differentiated product to another. With spatial competition, each firm has a local monopoly in the area immediately surrounding its establishment, but unrestricted entry leads to keen competition. Firms will continue to enter the market until economic profit drops to zero.

WHAT'S NEXT?

This introductory chapter sets the stage for the economic analysis of cities in the rest of the book. Here are some of the big questions we'll address in coming chapters:

- Why do cities exist?
- Are cities too big or too small?
- What causes urban economic growth?
- Why is employment in modern cities so widely dispersed?
- Why is there so much segregation with respect to race and income?
- Why do economists advocate a tax of about 7 cents per mile for all driving and about 27 cents per mile for driving on congested roads?
- Why do so few people take mass transit?
- Why is crime higher in cities?
- Why does the typical metropolitan area have dozens of municipalities?
In answering these and other questions, we will use the five axioms of urban economics. In addition, we will use a number of economic models to explore the spatial aspects of decision making. It’s worth noting that much of the analysis in the book reflects advances in urban economics in the last 10 to 15 years, in both theoretical modeling and empirical analysis. In fact, so much has grown that writing the sixth edition of the book involved an extreme makeover, with only a few paragraphs and figures surviving from the fifth edition.

REFERENCES AND ADDITIONAL READING


Appendix: Census Definitions

The U.S. Census Bureau has developed a variety of geographical definitions relevant to urban economics. Since much of the empirical work in urban economics is based on census data, a clear understanding of these definitions is important. This appendix provides the details of the census definitions.

URBAN POPULATION

The first three definitions deal with the urban population and are based on the census block, the smallest geographical unit in census data. A census block is defined as an area bounded on all sides by visible features (streets, streams, or tracks) or invisible features (property lines or political boundaries). The typical census block has between a few dozen and a few hundred residents. A block group is a group of contiguous census blocks. There are two types of urban areas:

1. Urbanized area. An urbanized area is a densely settled core of census block groups and surrounding census block groups that meet minimum population density requirements. In most cases, the density requirement is 1,000 people per square mile for the core block groups and 500 people per square mile for the surrounding blocks. Together, the densely settled blocks must encompass a population of at least 50,000 people. In 2000, there were 464 urbanized areas in the United States.

2. Urban clusters. An urban cluster is a scaled-down version of an urbanized area. The total population of the census blocks that make up an urban cluster is between 2,500 and 50,000 people. In 2000, there were 3,112 urban clusters in the United States.

3. Urban population. The Census Bureau defines the nation’s urban population as all people living in urbanized areas and urban clusters. Based on this definition, 79 percent of the population lived in urban areas in 2000.

METROPOLE AND MICROPOLITAN STATISTICAL AREAS

The Census Bureau has a long history of changing its definitions of metropolitan areas. The general idea is that a metropolitan area includes a core area with a substantial population nucleus, together with adjacent communities that are integrated, in an economic sense, with the core area. Over the years, the labels for metropolitan areas have changed from standard metropolitan area (SMA) in 1949, to standard metropolitan statistical area (SMSA) in 1959, to metropolitan statistical area (MSA) in 1983, to metropolitan area (MA) in 1990, which referred collectively to metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs—the largest metropolitan areas), and primary metropolitan statistical areas (PMSAs—parts of CMSAs).

The new label for areas considered metropolitan, implemented in 2000, is core based statistical area (CBSA). Each CBSA contains at least one urban area (either an urbanized area or an urban cluster) with at least 10,000 people and is designated as either a metropolitan area or a micropolitan area.

1. Metropolitan area. A metropolitan area includes at least one urbanized area with at least 50,000 people.

2. Micropolitan area. A micropolitan area includes at least one urban cluster of between 10,000 and 50,000 people.

In 2000, there were 361 metropolitan statistical areas and 559 micropolitan statistical areas in the United States.

The building blocks for metropolitan and micropolitan areas are counties. For a particular CBSA, central counties are ones in which at least 5,000 people or 50 percent of the population resides within urban areas with at least 10,000 people. Additional outlying counties are included in the CBSA if they meet minimum thresholds of commuting rates to or from the central counties. Specifically, at least 25 percent of workers in an outlying county must work in one of the central counties, or at least 25 percent of the jobs in an outlying county must be filled by residents of one of the central counties.

Together CSBAs contain 93 percent of the nation’s population, with 83 percent in metropolitan areas and 10 percent in the smaller micropolitan areas. The percentage of the population in CSBAs (93 percent) exceeds the percentage in urban areas (79 percent) because CSBAs encompass entire counties, including areas outside urban areas (defined by the smallest geographical unit, the census block).
PRINCIPAL CITY

The largest municipality in each metropolitan or micropolitan statistical area is designated a principal city. Additional cities qualify as "principal" if they meet minimum requirements for population size (at least 250,000 people) and employment (at least 100,000 workers). The title of each metropolitan or micropolitan statistical area consists of the names of up to three of its principal cities and the name of each state into which the metropolitan or micropolitan statistical area extends. For example, the name for Minneapolis metropolitan area is Minneapolis-St. Paul-Bloomington, MN-WI, indicating that it includes parts of two states with two other municipalities large enough to merit listing. For most metropolitan areas, the label includes only one principal city. About a dozen large metropolitan areas are divided into smaller groupings of counties called metropolitan divisions.

PART ONE

Market Forces in the Development of Cities

In a market economy, individuals exchange their labor for wage income, which they use to buy consumer goods and services. How do these market transactions affect cities? As we'll see in Chapter 2, cities exist because of the benefits of centralized production and exchange. We'll look at the rationale for the development of cities based on trade, production, and processing raw materials. Chapter 3 explores agglomeration economies, the economic forces that cause firms to cluster in cities to share the suppliers of intermediate inputs, share a labor pool, get better skills matches between workers and firms, and share knowledge. Chapter 4 explores the economic forces behind the development of cities of different size and scope. We'll look at how worker utility varies with city size and see why the equilibrium city size often exceeds the optimum size. Chapter 5 explores the sources of urban economic growth (increases in per-capita income) and urban employment growth. It also addresses the question of who benefits from employment growth and describes some of the techniques used by economists to predict future employment growth.
CHAPTER 2

Why Do Cities Exist?

Nobody ever saw a dog make a fair and deliberate exchange of one bone for another with another dog. —Adam Smith

Cities exist because individuals are not self-sufficient. If each of us could produce everything we consumed and didn’t want much company, there would be no reason to live in dirty, noisy, crowded cities. We aren’t self-sufficient, but instead specialize in a labor task—writing software, playing the accordion, performing brain surgery—and use our earnings to buy the things we don’t produce ourselves. We do this because labor specialization and large-scale production allow us to produce and consume more stuff. As we’ll see in this chapter, production happens in cities, so that’s where most of us live and work. By living and working in cities, we achieve a higher standard of living but put up with more congestion, noise, and pollution.

To explain why cities exist, we’ll start with a model that implies that they don’t. In the model of backyard production, every consumer is a producer, and all production occurs in backyards (or apartment roofs). In other words there is no need for concentrated production or population. As we drop the assumptions of the backyard production model, the new models imply that cities will develop. In other words, the short list of assumptions in the model identifies the key factors behind the development of cities.

A REGION WITHOUT CITIES—BACKYARD PRODUCTION

Consider a region that produces and consumes two products, bread and shirts. People use the raw materials from land (wool and wheat) to produce the two consumer products. The following assumptions eliminate the possibility of cities.

- Equal productivity. All land is equally productive in producing wheat and wool, and all workers are equally productive in producing shirts and bread.
• Constant returns to scale in exchange. The unit cost of exchange (the cost of executing one transaction, including transportation cost) is constant, regardless of how much is exchanged.

• Constant returns to scale in production. The quantity of shirts produced per hour is constant, regardless of how many shirts a worker produces. The same is true for bread production.

Together these assumptions eliminate the possibility of exchange and guarantee that each household will be self-sufficient. If a person were to specialize in bread and then trade some bread for shirts, she would incur a transaction cost equal to the time required to execute the trade. Under the assumption of equal productivity, there is no benefit from specialization because everyone is equally productive. Under the assumption of constant returns to scale, there is no benefit from producing shirts in factories because an individual is just as efficient as a shirt factory. In sum, exchange has costs without any benefits, so every household will be self-sufficient, producing everything it consumes.

The absence of exchange guarantees a uniform distribution of population. If population were concentrated at some location, competition for land would bid up its price. People in the city would pay a higher price for land without any compensating benefit, so they would have an incentive to leave the city. In the locational equilibrium, the price of land would be the same at all locations, and population density would be uniform. Recall the first axiom of urban economics:

Prices adjust to ensure locational equilibrium

In this case, all sites are equally attractive, so locational equilibrium requires the same price of land at all locations.

A TRADING CITY

Now that we have a short list of assumptions under which cities don’t develop, let’s drop the assumptions, one by one, and see what happens. We’ll start by dropping the assumption of equal productivity for all workers. Suppose households in the North are more productive in producing both bread and shirts. This could result from differences in soil conditions, climate, or worker skills. Table 2–1 shows the output per hour for the two regions. While each worker in the South can produce one shirt or one loaf per hour, workers in the North are twice as productive in producing bread and six times as productive in producing shirts.

<table>
<thead>
<tr>
<th>TABLE 2–1 Comparative Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
</tr>
<tr>
<td>Bread</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Output per hour</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>Opportunity cost</td>
</tr>
<tr>
<td>3 shirts</td>
</tr>
</tbody>
</table>

Comparative Advantage and Trade

A region has a comparative advantage in producing a particular product if it has a lower opportunity cost. For every shirt produced, the North sacrifices 1/3 loaf of bread, so that's the opportunity cost of a shirt. In the South, the opportunity cost of a shirt is one loaf. The North has a lower opportunity cost for shirts, so it has a comparative advantage in producing shirts. It is sensible for the North to specialize in shirts (and not produce any bread) because, although the North is twice as productive as the South in producing bread, the North is six times as productive in producing shirts.

Comparative advantage may lead to specialization and trade. Suppose the two regions are initially self-sufficient, with each household producing all the bread and shirts it consumes. Table 2–2 shows what happens if a North household switches one hour from bread to shirt production, and a South household goes in the other direction, switching two hours from shirt to bread production. The first row shows the changes in production: –2 loaves and +6 shirts for North; +2 loaves and –2 shirts for South. As shown in the second and third rows, if the households exchange two loaves and four shirts, each has a gain from trade of two shirts. After specialization and exchange, each household has just as much bread as before and two additional shirts.

What about transaction costs? The transaction cost is the opportunity cost of the time required to exchange products and is equal to the amount of output that could be produced during that time. For example, a North household can produce six shirts per hour, so the opportunity cost for a 10-minute (1/6 hour) transaction is one shirt. As long as the transaction time is less than 1/3 hour (two shirts), trade is beneficial. The South household, with a lower opportunity cost, has a lower transaction cost.

Scale Economies in Exchange

The presence of specialization and trade will not necessarily cause a city to develop. The second assumption of the backyard-production model is that there are constant returns to scale in exchange. Under this assumption, an individual household is just as efficient in executing trades as a trading firm, so there is no reason to pay a firm to execute an exchange. Therefore, each North household will link up with a South household to exchange shirts and bread directly, without intermediaries.

<table>
<thead>
<tr>
<th>TABLE 2–2 Specialization and Gains from Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
</tr>
<tr>
<td>Bread</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Change in production from specialization</td>
</tr>
<tr>
<td>–2</td>
</tr>
<tr>
<td>Exchange 4 shirts for 2 loaves</td>
</tr>
<tr>
<td>+2</td>
</tr>
<tr>
<td>Gain from trade</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Trading firms will emerge if there are economies of scale associated with exchange and trade. Recall the fourth axiom of urban economics:

Production is subject to economies of scale

A trading firm could use indivisible inputs such as a large truck to transport output between North and South. Similarly, workers who specialize in transportation tasks will be more efficient than workers who spend most of their time producing bread or shirts. In general, because trading firms have lower transaction costs, individual households will pay trading firms to handle exchanges.

The emergence of trading firms will cause the development of a trading city. To fully exploit scale economies, trading firms will locate at places that can efficiently collect and distribute large volumes of output. The concentration of trade workers will bid up the price of land near crossroads, river junctions, and ports. The increase in the price of land will cause people to economize on land by occupying smaller residential lots. The result is a place with a relatively high population density—a city.

**TRADING CITIES IN URBAN HISTORY**

Our simple model of the trading city suggests that trading cities develop when comparative advantage is combined with scale economies in transport and exchange. This observation provides some important insights into the history of cities before the Industrial Revolution of the 1800s. Most of the workers in these trading cities didn’t produce goods, but instead collected and distributed goods produced elsewhere, such as agricultural products from the hinterlands and handcrafted goods from various locations. Trade was a risky business, and firms in the trading city provided insurance, credit, investment opportunities, banking, and legal services.

**Trading Cities in World History**

Trading cities have a long history. In the third millennium B.C., Phoenicians used fast sailing ships to serve as traders for the entire Mediterranean basin, trading dye, raw materials, foodstuffs, textiles, and jewelry. They established trading cities along the Mediterranean coast in present-day Lebanon. Around 500 B.C., Athens was a thriving site for regional trade, exchanging household crafts and olive products for food and raw materials from the countryside. During the 11th and 12th centuries, Italian city-states forged agreements with the Byzantine and Islamic rulers for trade with North Africa and the East. The Europeans traded wood, iron, grain, wine, and wool cloth for medicines, dyes, linen, cotton, leather, and precious metals. This trade was the major force behind the growth of Venice, Genoa, and Pisa.

Some cities were built on coercive transfer payments rather than voluntary trade. The Athenian empire developed in the aftermath of the successful war against Persia in the fifth century B.C. After the Greek city-states repelled the Persian invasion, they formed the Delian League for joint defense and later to carry the war into Asia Minor. By the end of the successful campaign, Athens controlled the league and transformed the voluntary contributions of member city-states into payments of tribute to Athens. The system of homage and tribute led to the Peloponnesian War between the Athenian Empire and Sparta (431 to 404 B.C.). The war ended when Athens renounced control over its empire and demolished its defensive walls.

By the third century A.D., Rome had a population exceeding 1 million. The Romans established colonial cities throughout Europe and focused on collecting the agricultural surplus while they neglected urban production activity (Hohenberg and Lees, 1985). Instead of exchanging urban goods for agricultural products, Rome used conquest and tribute to feed its population. In the fourth and fifth centuries, attacks from Germanic tribes disrupted the Roman collection system. It appears that there was little interest outside of Rome in restoring the “trade” routes, so the losses from successive attacks were cumulative. If Rome had relied to a greater extent on voluntary exchange, the colonies would have had a greater stake in maintaining the exchange network and the Western empire might have recovered from the Germanic raids.

What are the lessons from the rise and fall of Athens and Rome? Early in its history, Athens thrived under a system of voluntary trade with other areas, exchanging urban goods for food from the countryside. The Athenians eventually switched to a system of conquest and tribute, resulting in war and the decline of the city. Mumford (1961) suggests that the city of Rome should have been called "Panapaopolis" to indicate its dependence on the labors of outsiders. The decline of Rome was caused in part by the disruption of its collection system by the Germanic raids. Perhaps the lesson is that cities based on coercive transfer payments are not sustainable.

**Trading Cities in American History**

The history of urban America illustrates the role of transport costs and comparative advantage in trading cities (Bartlett, 1998). In the 1700s, most cities served largely as trading posts for ocean trade. On the eastern seaboard, cities collected agricultural products from their hinterlands to the west and shipped them overseas. The volume of trade was limited by the dirt roads serving the interior. Travel was always slow and, in times of rain and melting snow, slipper. The Pennsylvania Turnpike, built with stone and gravel in 1792, increased travel speeds to a steady two miles per hour, increasing the market area and trading volume of the city of Philadelphia.

Farther to the north, New York State took more drastic steps, completing the 360-mile Erie Canal in 1825. The canal linked New York City, with its natural harbor, to vast agricultural areas to the north and west, and it cut freight costs from about 20 cents per ton mile to 1.5 cents. An additional canal connecting Lake Champlain to the Hudson River extended the market area of New York City to northern New England. The vast transportation network increased the volume of trade through New York City, increasing its size. By 1850, the city had a population of half a million, about 20 times its size at the end of the American Revolution. Other cities, including competitors to the south (Baltimore and Philadelphia),
responded by building canals to connect hinterlands and ports, and by 1845 there were over 5,300 miles of artificial waterways in the United States.

Comparative advantage also plays a role in urban history. Eli Whitney’s cotton gin (1794) provided a means of removing the sticky seeds of green-seed cotton, which could be grown in the south. The total output of cotton increased by a factor of 50 over a 15-year period, with most of the output going to inland areas far from the east coast ports. American cotton was transported along rivers to New Orleans for shipment to textile firms in New England and Europe. The increase in cotton trade caused the rapid growth of New Orleans at the mouth of the Ohio River. Enclaves such as Mobile, Alabama and Natchez, Mississippi, and the development of upriver commercial cities such as Mobile, Alabama and Natchez, Mississippi.

Later innovations in transportation reduced transport costs and contributed to the development of trading cities. Before the introduction of the steamboat in 1807, traffic was strictly downstream: After cargo was unloaded at the terminal point, wooden boats were broken up for lumber. The steamboat allowed two-way traffic and cut river freight costs, increasing the volume of trade and the size of river cities. Later, the steam engine was used to power locomotives, and railroad freight replaced river shipping as the principal means of transporting goods. Between 1820 and 1890, the ratio of railroad freight to river freight went from 0.10 to 2.0, and the volume of railroad freight increased by a factor of 240. The shift from river to railroads caused the decline of commercial cities along rivers and the rise of cities along the vast railroad network.

A FACTORY TOWN

The third assumption of the backyard-production model is constant returns to scale in production. We’ll maintain this assumption for bread production, but apply the fourth axiom of urban economics to shirt production:

Production is subject to economies of scale

A shirt factory will use indivisible inputs (machines) and allow workers to specialize in narrowly defined tasks, leading to a higher output per worker and lower average cost. Suppose a household can produce either a loaf of bread or one shirt per hour. A worker in a shirt factory is six times as productive as a home worker, so the factory worker produces six shirts per hour.

Determining Wages and Prices

We assume that workers are perfectly mobile, so the utility level of a city worker must be the same as the utility level for a rural worker. Recall the first axiom of urban economics:

Prices adjust to ensure locational equilibrium

A factory must pay its workers enough to make them indifferent between working in the factory town and in the rural area. A rural worker earns one loaf of bread per hour, so city workers must earn one loaf per hour plus an amount high enough to offset the higher cost of living in the factory town, such as higher land prices. For example, if the cost of urban living is 50 percent higher, locational indifference requires an hourly wage of 3/2 loaves of bread. A city worker will pay 1/2 loaf for land, leaving one loaf per hour of factory work, the same that she could earn producing bread in a rural area.

What’s the price of factory shirts? The price must be high enough to cover the costs of labor and the indivisible inputs used to produce shirts. In Table 2–3 the labor cost per hour is the wage (3/2 loaves) and the hourly cost of indivisible inputs is 1/2 loaf. Adding these together, the hourly cost of producing shirts is two loaves of bread. To translate this into a cost per shirt, recall that a factory worker produces six shirts per hour, so the cost per shirt is one-sixth of the cost per hour, or 2/6 = 1/3 loaf. Therefore, for zero economic profit, the price per shirt must be 1/3 loaf of bread.

Suppose there is a single shirt factory in the region. The factory competes with homemade shirts, and will sell shirts to any household for which the net price of factory shirts is less than the cost of a homemade shirt. The cost of a homemade shirt is the one loaf of bread that is sacrificed to produce a shirt. The net price of a factory shirt equals the price charged by the factory (1/3 loaf) plus the opportunity cost of travel to and from the factory to buy the shirt.

The Market Area of a Factory Town

Figure 2–1 (page 24) shows the net price of factory shirts and the market area of the shirt factory. As shown by point f, the net price for a consumer located just across the road from the factory (distance = 0) is the factory price, equal to 1/3 = 4/12 loaf of bread. Other consumers bear a travel cost when they buy factory shirts, so the net price is higher. Suppose the travel time is 1/12 hour per round-trip mile. It takes 1/12 hour to complete a round trip of one mile in each direction. In an hour, a rural household can produce one loaf of bread, so in 1/12 of an hour of travel, it sacrifices 1/12 loaf. For example, at point g (two miles from the factory), the net price of a factory shirt is 6/12 loaves, equal to 4/12 paid at the factory plus 2/12 in travel cost (forgone bread production at home).

The market area of the factory is the area over which it underprices the home production of shirts. In Figure 2–1, the horizontal line shows the opportunity cost of homemade shirts, which is one loaf of bread. The net price of factory shirts is 4/12 at the factory and increases by 1/12 per mile, reaching one loaf at a distance of eight.
in industrial cities. In contrast to the earlier trading cities, workers in factory cities produced products rather than simply distributing products produced elsewhere.

**Innovations in Manufacturing**

One of the key innovations of the Industrial Revolution was Eli Whitney’s system of interchangeable parts for manufacturing, developed around 1800. Under the traditional craftsman approach, the component parts of a particular product were made individually—and imprecisely. Skilled craftsmen were necessary to produce the parts and then fit them all together. Under Whitney’s system, the producer made a large batch of each part, using precise machine tools to generate identical parts. The identical parts were interchangeable, so unskilled workers could be quickly trained to assemble the parts. The replacement of handcraft production with standardized production generated large scale economies, causing the development of factories and factory cities.

Whitney applied this system to the production of muskets for the army. To prove to President-elect Jefferson and other government officials that his system would work without unskilled labor, he unloaded a random collection of parts onto the floor and had the officials assemble the muskets. He got the contract to manufacture 10,000 muskets and built a factory in New Haven, Connecticut, close to a stream that he used to power the factory. His system, which became known as the American System of Manufacturing, became the standard system for mass production.

The new system of manufacturing caused the development of factory cities. New machines, made of iron instead of wood, were developed to fabricate products in large factories. Manual production by skilled artisans was replaced by mechanized production using interchangeable parts, specialized labor, and steam-powered machines. Mass production decreased the relative cost of factory goods, causing the centralization of production and employment in large industrial cities.

As an illustration of the role of scale economies in the development of cities, consider the sewing machine, which was developed in the middle of the 19th century. At the beginning of the century, about four-fifths of the clothing worn in the United States was hand-sewn in the home for members of the household, and the rest was hand-sown by tailors. The sewing machine (patented in 1846) allowed factories to underprice home producers, and by 1890 nine-tenths of U.S. clothing was being made in factories. New cities developed around the clothing factories.

A similar story line applies to shoes. Before 1700, most shoes were produced in the home or the local village. The cost of transportation was so high that local production was efficient. Over time, transportation costs decreased, and the putting-out system was implemented in the 1700s: Shoe producers distributed raw materials to cottage workers, collected their rough output, and finished the shoes in a central shop. As new shoemaking machines were developed, the number of operations performed in the central shops increased. The McKay sewing machine (for which a patent was granted to Lyman Blake in 1858) mechanized the process of sewing the soles to the uppers. The scale economies in shoe production increased to the point that shops became genuine factories, and cities developed around the shoe factories.

**THE INDUSTRIAL REVOLUTION AND FACTORY CITIES**

Our simple model of the factory city suggests that a factory town develops because scale economies make factory shirts cheaper than homemade shirts. The Industrial Revolution of the 19th century produced innovations in manufacturing and transportation that shifted production from the home and the small shop to large factories.
Chapter 2  Why Do Cities Exist?

Innovations in Transportation
Innovations in intercity transportation contributed to industrialization and urbanization. As we saw earlier in the chapter, the dirt roads of the 1700s were replaced by turnpikes, and the construction of canals allowed a more dense network of inland water transport. The development of the steamship allowed two-way travel on major rivers, and the railroad system increased the speed and reach of the transportation system. All of these innovations decreased the relative price of factory goods, contributing to the growth of factory cities.

Innovations in Agriculture
One of the three conditions for the development of cities is an agricultural surplus to feed city dwellers. The Industrial Revolution generated a number of innovations that increased agricultural productivity. Farmers substituted machinery for muscle power and simple tools, increasing the output per farmer. The increased agricultural productivity freed people to work in urban factories and commercial firms. Between 1800 and 1900, the share of the population living in cities increased from 6 percent to 35 percent, reflecting the decrease in the number of agricultural workers required to feed city dwellers.

Consider first the sowing side of agriculture. At the start of the 19th century, plows were fragile, awkward, and often made of wood. These inefficient plows were replaced in the 1830s by the cast-iron plow, which was produced in factories in Pittsburgh and Worcester. In the 1840s, John Deere introduced the steel plow, which was lighter, stronger, and easier to handle. Later innovations allowed the farmer to adjust the depth and angle of the plow blade, increasing productivity further.

Consider next the reaping side of agriculture. In 1831, McCormick combined several earlier innovations into a horse-drawn harvesting machine that increased the productivity of the most labor-intensive part of agriculture. Using a horse-drawn reaper, two people could harvest the same amount of grain as eight people using traditional harvesting methods.

Other innovations contributed to higher agricultural productivity. The development of agricultural science led to innovations in planning, growing, harvesting, and processing. Innovations in transportation cut transport costs and allowed each farmer to serve a wider market area. Because of rising productivity, the share of employment in agriculture decreased over the 19th and 20th centuries, from over 90 percent to less than 3 percent.

Energy Technology and Location Decisions
During the Industrial Revolution, the location pattern of factory towns reflected changes in energy technology. The first factories used waterwheels turned by waterfalls and fast-moving streams to translate moving water into mechanical motion. The power was transmitted by systems of belts and gears. Textile manufacturers built factories along backcountry streams in New England and used waterwheels to run their machines. Some examples of waterwheel cities are Lowell, Lawrence, Holyoke, and Lewiston.

The refinement of the steam engine in the second half of the 19th century made energy a transportable input. A key innovation was John McNaught's development of a compounding engine (using steam twice, at descending pressures, to drive pistons) in 1845. The steam engine could be operated anywhere, with the only constraint being the availability of coal to fuel the engine. Some energy-intensive manufacturers located near the coal mines in Pennsylvania. Others located along navigable waterways and shipped coal from the mines to their factories. In New England, textile firms shifted from backcountry waterfall sites to locations along navigable waterways. Production shifted to the Fall River–New Bedford area along the south coast of New England. The later development of the railroad gave coal users another transport option, causing the development of factories along the vast network of rail lines. In general, the steam engine widened the location options for factories.

The development of electricity changed the location patterns of factories. Electricity generators were refined in the 1860s, and the electric motor was developed in 1888. Factories replaced belt-and-gear systems driven by a central steam engine with small electric motors for individual machines. The first factory to use electric power was adjacent to a hydroelectric generating facility at Niagara Falls. Rapid improvements in the electricity transmission soon allowed factories to be hundreds of miles from hydroelectric and coal-powered generating plants. Between 1900 and 1920, the share of factory horsepower from electric motors increased from 2 percent to 33 percent.

The development of electricity made factories more flexible. A firm could tap water power without locating close to the stream and use coal without shipping the bulky fuel to the factory. In general, the development of electricity decreased the importance of energy considerations in location decisions, causing firms to base their location choices on the accessibility to other inputs and to consumers.

A SYSTEM OF FACTORY TOWNS
We can widen our horizon by looking at the entire region and consider the possibility of additional factory towns. Firms can enter the shirt industry by building shirt factories at different locations, and each firm will have a local monopoly in the area surrounding its factory. Recall the fifth axiom of urban economics:

**Competition generates zero economic profit**

If there are no restrictions on entry, firms will continue to enter the market until economic profit is zero.

Figure 2–2 (page 28) shows the equilibrium in the region. The horizontal axis measures distance from a coastline. The rectangular region is 48 miles wide, and in equilibrium it has three shirt factories, each with a market area 16 miles wide.
market areas of the factories span the region: Every location in the region lies within the market area of some factory. There is complete labor specialization: Workers in factory towns produce shirts (and receive bread as wages), and workers in rural areas produce bread (and pay bread to get factory shirts).

This is an equilibrium because each firm makes zero economic profit and workers are indifferent between rural and city life:

- **Zero economic profit.** The factory price of 4/12 loaf equals the average cost of producing shirts, including the cost of urban workers and the cost of indivisible inputs.

- **Locational indifference for workers.** The wage for factory workers is high enough to cover (1) the opportunity cost of working in factories rather than producing bread in the rural area and (2) the higher cost of urban living (land rent).

What about rural residents? For a rural resident just outside the factory towns, the net price of a factory shirt is 4/12 loaf of bread, compared to a homemade cost of 1 loaf. At the other extreme, a rural household eight miles from the factory pays a net price of one loaf per shirt (4/12 + 8/12 in travel cost). Recall the first axiom of urban economics:

**Prices adjust to ensure locational equilibrium**

In this case, the price of land in rural areas will adjust to make people indifferent between locations that differ in their accessibility to the shirt factory. The shorter the distance to the factory, the lower the net price of factory shirts, and the more a household is willing to pay for land. In other words, the price of land adjusts to fully compensate for differences in accessibility.

Landowners benefit from the scale economies in production that generate the regional system of factory towns. In the rural areas, the price of land is higher at locations close to the factory city. In the factory city itself, competition among workers for locations near the factory bids up the price of land.

**MATERIALS-ORIENTED FIRMS AND PROCESSING TOWNS**

Up to this point, we have ignored the cost of transporting the raw materials required to produce urban goods (shirts). We have implicitly assumed that factory workers harvest wool from wild sheep who wander by the factory at just the right time to be sheared for shirts. In the language of urban economics, we have assumed that the raw materials required for production are ubiquitous—available at all locations at the same price. This is an extreme case of a market-oriented industry, defined as an industry in which the cost of transporting output is large relative to the cost of transporting inputs.

Consider the opposite extreme. Suppose the materials are raw agricultural inputs, but output can be transported at zero cost. This is the extreme case of a materials-oriented industry, defined as an industry for which the cost of transporting material inputs is large relative to the cost of transporting output. For example, the sugar content of sugar beets is roughly 15 percent, so it takes seven tons of beets to produce one ton of sugar. Beet-sugar firms locate their plants close to the beet fields to economize on transport costs (Holmes and Stevenson, 2004). Beet-sugar producers will cluster in the regions of the country where weather and soil conditions are favorable for the production of sugar beets.

**Scale Economies and Market Areas**

The process of transforming sugar beets into sugar is subject to scale economies. Processors use indivisible inputs and engage in factor substitution, so the average processing costs decrease as the quantity increases. The typical sugar-beet processing plant employs 186 workers, about four times the average number of employees per plant in manufacturing.

The market area of a processing plant is determined by the net price farmers receive. The net price equals the price paid by the processor minus the cost of transporting the beets from the farm to the processing plant. In Figure 2–3, the horizontal
Competition generates zero economic profit

If there are no restrictions on entry, firms will continue to enter the market until economic profit is zero.

System of Processing Towns

The location of sugar-beet processing plants leads to the development of a system of processing towns. The people who work in the processing plants live nearby to economize on commuting, resulting in a place with a relatively high population density: a small city or town. As Holmes and Stevens (2004) show, beet-sugar processing plants locate in the regions where beet production occurs, and they carve out input market areas within each beet-growing region.

Note the similarities of the beet-sugar industry and the shirt industry. In the shirt industry, with relatively high cost of transporting output, each firm gets a local monopoly, with all consumers patronizing the nearest factory. In the beet-sugar industry, with relatively high cost of transporting input, each firm gets a local monopoly, with all farmers selling their output to the nearest beet-sugar plant.

Other Examples of Materials-Oriented Industries

The same logic applies to other materials-oriented industries (Kim, 1999). The production of leather requires hides and tannin (from tree bark) for the tanning process. The tannin content of bark is only 10 percent, so it takes a lot of bark to produce a ton of leather. In 1900, U.S. leather producers located close to forests to economize on the transport costs of tannin. In the 20th century, improvements in the extraction process and the development of synthetic tannin reduced the orientation toward forest sites, and firms moved closer to other input sources.

The location decisions of steel producers reflected changes in the input requirements of coal and iron ore. Early in the history of the industry, a ton of steel required five tons of coal and two tons of ore, and steel production was concentrated near coal deposits. Technological innovations reduced the coal content, and steel producers were pulled toward locations that provided access to ore deposits, including sites on the Great Lakes that offered water access to ore from the Masabi Range in Minnesota. Each ton of steel required 175 tons of water, and the Great Lakes sites also provided a plentiful supply of water.

Chapter 2 Why Do Cities Exist?

The location decisions of these and other materials-oriented industries caused the development of processing towns. Leather towns developed around tanneries and steel towns developed around steel mills. Only producers of these products locate near forests, causing the development of lumber towns centered on sawmills. Ore processors locate near mineral deposits, causing the development of mining towns.

SUMMARY

Cities exist because of the benefits of centralized exchange (trading cities) and centralized production (factory and processing cities). We have focused on the market forces that generate cities. For a discussion of other possible reasons for cities, such as religion and defense and their role in the development of the first cities, see Munn (1961) and O'Sullivan (2005). Here are the main points of this chapter:

1. A trading city develops when comparative advantage is combined with scale economies in exchange.
2. A factory-city develops when there are scale economies in production.
3. The Industrial Revolution caused massive urbanization because of its innovations in agriculture, transportation, and production.
4. Changes in energy technology altered the location decisions of firms, with water power generating factories along streams, steam power generating factories along rivers and railroads, and electricity making firms more footloose.
5. Spatial competition among firms generates a market area for each firm and a system of cities.

APPLYING THE CONCEPTS

1. Matter Transmitter in Trading City

Consider a region with two standardized products (bread and shirts) and a single trading city. Transport is initially by horse-drawn wagons. Predict the effects of a new matter transmitter, which can instantly transport goods (but not people), with a zero marginal cost of transport. The transmitter is an indivisible input, and it is expensive.

a. Predict the effects of the transmitter on (i) the volume of trade and (ii) the population of the trading city. Will the matter transmitter cause the trading city to grow, shrink, or disappear?

b. How would your answers to (a) change if the matter transmitter were relatively cheap?

2. Self-Singing, Choral Groups, and the Internet

Consider a region where households produce and consume two products, bread and live musical performances. All workers are equally productive at producing bread and music. The production of bread is subject to constant returns to scale, with one hour required to produce each loaf. In an hour, a single person can produce one unit of music for herself. A choral group of 20 people working for an hour (57 minutes of practice and 3 minutes of performance) can produce
one unit of music for an audience of 80 people. Assume that the opportunity cost of actually listening to the music is zero.

a. What is the appropriate price for choral music, that is, the price (in bread) paid by each person who listens?

b. If the travel cost is 1/8 hour per round-trip mile, what is the market area of the choral group? Defend your answer with a completely labeled graph.

c. Suppose choral music becomes available on the Internet, and the provider charges 1/2 loaf per song. What is the new market area of the choral group? Illustrate with a completely labeled graph.

3. Spring-Loaded Sneakers

Using Figure 2–1 as a starting point, suppose all consumers switch to spring-loaded sneakers, decreasing walking time per round-trip mile from 1/12 to 1/18 hours.

a. Depict graphically the effects of the new sneakers on the market area of the shirt factory.

b. Modify Figure 2–2 to generate a region of factory cities with market areas that span the entire region (48 miles wide). Do the new sneakers increase or decrease the number of cities?

4. Beer and Wine

Most breweries locate close to their customers (far from their primary input sources), while most wineries locate close to their input sources (far from their consumers).

a. Why?

b. Consider a region that is 120 miles wide. Beer consumers are uniformly distributed throughout the region, while grapes are uniformly distributed through the western half of the region. There are two evenly spaced wineries and two evenly spaced breweries. Where do they locate? Illustrate with a graph.

5. Helicopters and Cities in Retireland

Consider Retireland, where no one works and everyone consumes a single good (food), which is imported from another region and can be purchased (one meal at a time) from the nearest vending machine. Alternatively, food can be delivered by helicopter, by dropping a meal through a food slot on the roof. The price of a delivered helicopter meal is $6 and the price of a vending-machine meal is $2. The travel cost is $0.04 per round-trip meter ($0.02 per meter traveled).

a. What is the market area of a vending machine? Illustrate with a graph.

b. How does the price of land vary within the region?

c. What are the implications for population density and cities?

REFERENCES AND ADDITIONAL READING


3. Uncertainty, satisfying, and information are important elements of the behavioral approach to location decisions. What is their significance? How are they handled by the profit maximization, institutional, and behavioral approaches?

4. What is involved in determining the comparative advantage of a region with respect to the types of industrial structure found active in the region?

5. How do localization and urbanization economics differ? What specific aspects of profit maximization relate to localization and urbanization economics?

6. How does a pecuniary externality help explain firm location?

7. Describe the various elements of Porter's diamond of regional competitiveness.

8. Do different types of location decisions (start-ups, expansions, etc.) rely on different location factors? If so, how do they differ?

9. A community can influence some location decisions. Which ones, and how might the community exert that influence?

10. Describe clusters and how they affect community economic development.

NOTES

1. By construction, the average costs curves used here include returns to land, labor, and capital. In the business accounting world, returns to land and capital are sometimes called normal profit. Economic profit exists when long-run average revenue exceeds long-run average costs. Economic and normal profits are distinct and separate ideas. In a competitive world, normal profits are positive, but economic profits are zero. Economists are sometimes referred to as excessive profit.

2. Decision-makers have a tendency to repeat prior successful decisions. For communities, this means that businesses continue to use the same production technique or location rather than trying new ones (Lloyd and Dicken 1977; Rees 1974).Townroe (1974) described three management conditions associated with location decisions: (1) lack of experience with the particular market and procedures to follow; (2) ignorance of all the relevant location possibilities; and (3) uncertainty about what decision criteria to use. Location decisions are made in a dynamic environment in which the firm and community affect each other. This learning and adaptation process is continuous. The adaptation may arise as altered production scheduling and processes, reduced sales and profits, or it may be relocation or closure. Townroe (1974) argued that the adjustment/adaptation process occurs because of unforeseen circumstances at the time of the initial decision, inadequate information was sought and used, critical factors for operation were not fully accounted for, and, finally, poor judgment was used initially.

3. This situation points out the asymmetry of information in the location decision process. The firm knows whether or not a subsidy is truly necessary to make the site either profitable or the site of first choice. The community does not know this information, thus there is a constant danger that it will oversubsidize the already profitable decision. Although the term regional is used, it is important to remember that community could also be used. Regional is used because we are generally referring to area-wide comparative advantage.

5. These primary factor resources are dealt with more completely in the second section of this text. As the astute reader will note, land, labor, and capital are only limiting in explaining regional comparative advantage. Later in the book we add to this list more latent primary factors of production that are increasingly important in explaining the locational decisions of firms and households. These latent factors include technology (Chapter 8), amenities (Chapter 9), and publicly provided goods and services (Chapter 10).

6. Clusters defy being classified according to some NAICS or SIC code. They are broader than a single NAICS category.

7. Typically, a firm seeks to change its location because it is experiencing some form of stress. This stress can be that the market has geographically moved, the source of input factors has geographically moved, or the firm faces a surge in demand for its output. A firm will generally seek many alternatives before changing location. These alternatives include expanding production by adding a second and third shift, expanding facilities on the present site, or partnering with another firm.

Community economies are dynamic entities that do not operate in isolation. They represent inextricably linked components that relate internal markets to the outside world. The vibrancy of a local community's economy can be thought of in terms of how effective its internal and external linkages are. This set of linkages provides the focus for this chapter and reflects the important aspects of the markets node of the Shaffer Star (Fig. 1.1).

We can think about community markets in two fashions. First, we can think of them in terms of goods and services markets and factor of production markets. Second, we can also think about internal and external markets. These latter markets can be goods and services or factors of production. In this chapter, we focus on the goods and services markets and defer the discussion of the factors of production markets to later chapters.

Market-oriented community economic development theories study the forces that affect the demand for the goods and services the community produces and how that is translated into community income or employment. If a community has a comparative cost advantage in the production and distribution (Chapter 3) of a good or service demanded in the external and/or internal market, then the community will attract the capital and labor necessary to produce the good or service. Over the long run, competitive market forces create an optimal spatial distribution of economic activity by selecting those production sites most profitable for the market served.

Local unemployment, low income, and slow growth represent short-run symptoms of a decline or shift in demand for the community's output. The market responds to this decline in demand by shifting capital and labor to more-productive uses within the community or in other communities. The presence of underutilized factors of production, such as unemployed labor, attracts other lines of production to the community. Persistent symptoms of local economic distress may indicate that the national economy may be improved if underutilized local factors of production relocate out of the community.

External markets are markets outside of the spatial boundaries of the community, while internal markets are markets inside the spatial boundaries of the community, which is that geographic area from which it draws the majority of its retail and service trade. Each of these represents a completely different theoretical perspective on how the economics of the community work. The external market is modeled by the export base theory. The internal market is modeled by central place theory and local market analysis.

We begin the chapter with the circular flow model of the local economy and then move on to export base theory, where we focus on how export or external markets are related to internal markets. Next, we narrow our discussion to internal markets and build on the ideas of the demand-maximizing firm outlined in Chapter 3. We do this because community economic development policy often overlooks the one market on which the community can often have the greatest impact: internal markets.

CIRCULAR FLOW

A useful organizational framework within which to better understand these economic linkages is the circular flow concept. This conceptual approach, which outlines how a community economy operates, is presented in its regionalized form in Figure 4.1. The general circular flow is made up of the basic actors involved in economic transactions and the markets within which supply and demand interact to discover prices and quantities. The circular flow diagram is linked to the outside world by imports/exports of goods and inflows/outflows of factor resources. Because we respect inflows and outflows of goods
resources are important determinants of regional linkage.

In our oversimplification of the world, the physical circular flow is represented by the arrows in Figure 4.1. Consider the linkage between households and the factor markets. Households own land, labor, and capital and offer it for sale (supply) in the factor markets. Firms demand these factors of production and pay for them in the forms of rent, wages, and interest. These payments represent income to households that, in turn, is used for consumption (and savings) in the goods and services market.

**EXTERNAL MARKETS:** EXPORT BASE THEORY

As in the circular flow model, there are internal as well as external markets in which the community functions. Community economic development policy has historically focused on only one small part of the circular flow representation, specifically the external market for goods and services. This focus has followed from a simple model of economic growth known as export base theory or economic base theory. In our broader view of the local economy in the circular flow model, narrowly focusing on exports limits the options available to the community. Export base theory, however, does provide a useful means of thinking about the community’s economy and how it changes.

Export base theory argues that the community’s economy can be divided into two sectors (Andrews 1979a, 1979b, 1979c; Blumenfeld 1955; North 1955, 1956; Tiebout 1956a, 1956b). The first sector is the export, or basic, sector. The export sector consists of that portion of the community’s goods and services market that trades with other areas. The export sector brings dollars into the community because someone outside the community purchases goods and services produced in the community. The second sector, termed the nonexport, nonbasic, or residualary sector, sells its product within the boundaries of the community (internal markets) and exists to support the export sector. Equation (4.1) displays the division of the total local economy into its basic and nonbasic sectors. It is a fairly simple theory claiming that the local economy can be divided into two parts: basic (or export) and nonbasic (or nonexport).

\[
\text{Total} = \text{Basic} + \text{Nonbasic} = E_b + E_{non}
\]  

(4.1)

The nonexport \((E_{non})\) component of the local economy \((E_b)\) is generally larger than the export sector \((E_b)\) but is dependent on the export sector. Thus, the export base model argues that any change in the export base leads to some multiple of the change in the total local economy. Specifically, the change in the export sector has a feedback or multiplier impact on the nonbasic sector.

This multiplier effect is represented graphically in Figure 4.2 (see below and Chapter 15 for a fuller discussion of multipliers). The basic sector \((E_b)\) shown on the horizontal axis drives the nonbasic

\[
\begin{align*}
\text{Slope of line represents the level of export influence} & \quad \beta = (E_b - E_{non})/E_b \\
rise \text{ change in } E_b & \quad \text{Y-intercept (a) reflects the level of economic activity present in region without exports} \\
\text{run change in } E_{non} &
\end{align*}
\]
sector ($\delta_3$) on the vertical axis. Any change in the basic sector has a two-part impact on the total economy. First is the initial change in the basic sector itself; the second is the impact on the supporting nonbasic sector. This second impact is some multiple of the change in the basic sector. The slope of the curve in Figure 4.2 maps out that multiplier relationship. A relatively flat multiplier relationship (small slope) suggests a weak relationship between the basic and nonbasic sectors of the local economy. A steep line (large slope) implies a strong relationship between the basic and nonbasic sectors.

Export base theory contends the development of a community depends on the vigor of its export or basic industries. The critical force in the community’s economic development is external demand, not the community’s ability to supply capital and labor or use technology. The timing and pace of the community’s economic development depend on the success of its export sector, the characteristics of the export sector, and the disposition of income received from export sales (North 1961).

The export sector carries external economic forces into the community. The characteristics of the export sector and the disposition of its income translate the external forces into community economic development. As we have seen, how changes in the export sector affect the community depends on the number and strength of linkages between the export and nonexport sectors. Furthermore, the distribution of income from the export sector and the ownership of resources used in the export sector assist in translating export sector changes into community economic development. For example, if the ownership of export base resources is external to the community, then changes in the export sector have a minimal impact on the community because the income is not recirculated or reinvested in the community. Finally, the availability of skills that permit the local labor force to work in the export sector also contribute to the success of translating external demand into local economic change.

The real criterion of the export function is whether the activity brings income into the community. Exports occur when an economic transaction occurs across the community’s economic boundaries. An export transaction occurs either as the movement of a good or service to the consumer or purchaser or as the movement of a good or service to the consumer or purchaser for the good or service (Andrews 1970b, 1970c). The movement of goods to the nonlocal consumer or purchaser is the most commonly perceived type of export transaction. Examples would include the shipment of agricultural, mining, and manufacturing goods to consumers located outside the community. The movement of the consumer to the good or service characterizes the second major type of export transaction. The movement of the consumer to services occurs when the nonresident consumer enters the community and consumes the service within the community. Some examples are recreation/tourism, regional medical services, and university education. In other words, the locally produced good or service can be shipped to the consumer or the consumer can travel into the community to make the purchase.

Historically, export base theory has focused on the top half of our circular flow model, the goods and services market. But the factor resources market can also be a source of exports. The owners of the factors of production are not limited to local markets. In the labor market, households could elect to commute to neighboring markets for employment and transport their wages back into the community. Capital can be invested in any number of markets outside the community; the returns on that investment represents an inflow of income into the community. Likewise, the ownership of nonlocal land yields returns to the community. For example, a retired farmer who moves to a warmer climate but continues to own and rent the farmland represents a source of new monies being injected into her new residence.

The volume of exports from a community can increase or decrease over time. An increase in export demand is caused when the external demand curve shifts right, income levels of nearby areas increase, comparative advantage in the community improves, or a community’s factor endowment increases. An example of the shift in external demand for some communities is the increased demand for microprocessors in the United States during the 1990s. A service with a high-income elasticity is the hospitality and recreation industry. The income levels of nearby areas are important because most communities sell more to nearby communities than to distant ones. The improvement of a community’s comparative advantage increases its exports because it permits lowering prices more than other communities can. Changes in technology or prices affect the profitability of using local resources to produce goods and services for export, except for the profitable exploitation of tar sands in northern Alberta, Canada, as crude oil prices increase and new technology becomes available. A change in tastes and preferences that increases the demand for specific products favorably impacts some communities because they have the resources, or factor endowment, permitting production of that good or service (see Chapter 9 on amenities for an example of this).

A community could also experience a decline in the volume of its exports over time. This could happen through depletion of its natural resource base or through a relative decline in its comparative advantage because the costs of land, labor, and capital increase, or because technological changes alter input combinations for which this community previously had an advantage. Or a leftward shift in external demand could be caused by a change in consumers’ tastes and preferences. If a community does not adjust to the forces decreasing its volume of exports, the community will find itself stranded outside the economic mainstream with a relatively even or absolutely worsening economic position. Just as communities experience increases or declines in their existing export base, they also experience the creation of new export sectors. The same factors causing an increase in exports can cause new export activity. Two additional forces of particular significance are governmental investments and new technology. Governmental investments in social overhead capital, such as water and sewer systems, industrial parks, and transportation systems, can eventually lead to new export businesses. New technology has a differential effect among communities. Some communities gain a competitive advantage through the early adoption of new technology. An example is the growth of the microprocessor industry and communities capable of supporting those businesses.

While widely used as a foundation for community economic development policy, export base theory requires several assumptions that frequently are not explicitly recognized (Pfister 1976; Richardson 1973; Tiebout 1956a). The simple linear relationship between basic and nonbasic sectors outlined in Figure 4.2 is built on a very strong set of assumptions:

1. Income and employment changes in a community depend upon changes in the level of exports, with no other stimulus for local change.
2. The factor propensities to consume locally, specifically the amount of local income spent for local products, is stable over time and over a relatively wide range of income change.
3. The amount of local income generated by each dollar of local spending does not change, that the local labor content does not vary over time.
4. There are no changes in the relative prices of capital or labor as their use increases or decreases (i.e., no shift from labor to capital or vice versa in response to changes in export demand).
5. The additional capital and labor required to expand production is available immediately and without any increase in wages or profits since the community has a perfectly elastic supply of capital and labor to meet increases in demand.
6. The economic structure of a community at one time will predict its future economic structure.
7. The homogeneous export sector implies that earnings from jobs and backwash linkages, among other factors in separate subsectors of the export sector, are roughly equivalent.
8. None of the local consumption of the goods and services sold for export comes from importing those goods and services (i.e., no cross-hauling).

From a policy perspective, these restrictive assumptions limit how useful export base theory is, but even more important are its internal inconsistencies from a purely theoretical perspective. First, all communities are exporting and none are importing. Who is purchasing the exports? Second, in the extreme, according to export base theory, it is impossible for the world economy to grow because there are no export markets. As in most of economics, extremely simple theories or ways of thinking about a problem often are partial and, when taken to a logical conclusion, lead to contradictory results.

Limitations

From the perspective of the community development economist, there are several major weaknesses in export base theory (Blumenfeld 1955; North 1955; Richardson 1969b; Tiebout 1956a, 1956b). First and foremost, export base theory is not a general theory of community economic development; it is more appropriate for smaller communities, simpler economies, and the about that for larger communities, complex economies, and the long run.
A theory that argues for a simple exogenous shift in export-demand as the source of economic development borders on naïvete; several other forces affect community economic development. The theory does not explain changes in the marginal propensity to consume locally as income changes. The export sector is not homogeneous, and changes in different parts of the export sector have disparate multiplier effects within the local economy. For example, think of the different linkages agriculture has with the local economy compared with manufacturing or the recreation industry. The changes in the export sector depend on nonexport or supporting economic activities that help create the comparative cost advantage for the export sector. For example, a change in export activities partially depends on the efficiency of the transportation system. In addition, export base theory is spatial and the role of market size and location is ignored. Nor does export base theory handle very well how changes in markets can alter the comparative advantages of the community. For example, a community endowed with forest resources may initially export lumber but changing markets suggest that a better use of the resources would be tourism. Export base theory lends no insights into how what markets might change.

Nonetheless, another advantage of the export base approach is its failure to explain community economic development occurring despite a decline in community exports. A community experiencing a decline in its exports could still grow because nonexport businesses grow enough to offset the decline. Improving the community’s terms of trade, such as the community getting a better deal for the goods it sells, means the community sells fewer goods yet still maintains or improves its total income.

Export base theory is only a partial theory. It fails to provide any insight into why new technologies are adopted or why institutions change over time. It does not explain how communities change because of changing comparative advantages, nor does it provide any insight into how communities are linked through both exports and imports. So for a system of communities, export base theory fails to explain the initial export changes, which are conveniently and justifiably assumed exogenous for the individual community.

The export base approach to community economic development does not account for structural changes in the community’s economy, which will forever depend on its export sector. As communities grow, they become less dependent on exports, and some nonbasic sectors become self-sufficient and maintain themselves without heavy reliance on export activities; which theory does not explain.

While export base theory argues that nonbasic activities are a passive component of the community’s economy, they need not be. In many cases, such activities such as downtown revitalization programs promote change without a change in the basic sector, which is in direct violation of the fundamental premise of export base theory. The source of growth can flow from the nonexport to the export sector, also in direct contradiction to export base theory. Some examples of this type of change follow:

1. The nonbasic sector becomes self-sustaining and is no longer dependent on the export sector because the local market has become large enough to support the business and there is sufficient trade among nonexport businesses and/or households (i.e., import substitution).

2. Local government investments change the amount and character of the social overhead capital base of the community and stimulate new export activities.

3. Non-market stimulated activity (e.g., stimulated by amenity resources and retirement) creates residuary activities to increase without an increase in export.

4. Improved efficiency in the nonbasic sector induces growth in the basic sector by making the export sector even more competitive relative to other areas.

5. The nonbasic sector experiences a higher rate of technological change than the basic sector. Here a shift of resources into the nonbasic sector increases community growth more than if the resources continue to be allocated to the export sector. Unlike example 4, the basic sector need not grow in this situation.

In summary, export base theory has intuitive appeal and simplicity. It also has a relatively sound theoretical foundation based on the concept that some local economic sectors transmit external economic forces into the community to stimulate further change. Changes in community income depend on changes in export demand. Exports increase because of a rightward shift in demand or an improved competitive position of the community, while exports decline because of a leftward shift in demand or a loss of competitive position. New export sectors appear in the community because of changes in tastes and preferences and in technology. Export base theory is more appropriate in smaller economic areas that are relatively more dependent on external trade. Likewise, it is more appropriate in simpler, less diverse economies. The importance of the export sector to the community declines with increases in the diversity and completeness of the local economy (i.e., self-sufficiency).

Export Base Multipliers

Historically, there are two approaches to economic accounting, or the means of organizing and reporting economic data: the Keynesian and Leontief approaches. It can be shown that the approaches represent two sides of the same coin. Both can be used to develop models of the economy and estimates of economic linkages and interrelationships. (We use the traditional Keynesian approach to income determination in this chapter and defer discussion of the Leontief approach to Chapter 15.)

In community economics, these linkages often are thought of through the multiplier effects. The multiplier measures the spending and responding of an exogenous injection of income (export income) that results in a total change in community income exceeding the original change. Rather than trace this spending and responding process for each change in exogenous income, multipliers provide a short cut.

Income (Y) is the sum of consumption, exports, imports, investment, and government expenditures:

\[ Y = C + X - M + I + G \]  

Total consumption (C) is driven by total income:

\[ C = a + CY \]  

Differentiating equation (4.9) with respect to exports (X) shows how a change in exports influences income. Specifically,

\[ \frac{dY}{dX} = \frac{1}{1 - (a - m)} \]  

where \( a \) (sometimes called a subsistence level of consumption) is some fixed level of consumption and \( m \) is the marginal propensity to consume.
where $Y_e$ is the change in basic sector income ($Y_e = X + I + G$) and $Y_{nab}$ is the change in nonbasic sector income ($Y_{nab} = C^*_g$).

The change in nonbasic income comes from the change in local consumption and its conversion into income:

$$Y_{nab} = C^*_g = Y_e(MPC)(PSY)$$  

(4.14)

where MPC is the proportion of income change spent locally (marginal propensity to consume locally) and PSY is the proportion of local consumption expenditures that becomes local income. Recognizing the income from expenditures process leads to refining the earlier multiplier formulation, equation (4.10), into

$$k = \frac{1}{1 - MPC \cdot PSY}$$  

(4.15)

The power of the Keynesian approach to thinking about economic linkages is its flexibility in defining linkages and capturing them via the multiplier. The export base multiplier is a function of local spending and its conversion into local income. More importantly, this discussion demonstrates the importance of local consumption, local government spending, and local investment to the conceptualization of economic linkages and the simple economic base multiplier. Furthermore, the importance of converting local spending into local income is recognized. Finally, the simple ratio of total to basic activity as an estimate of the multiplier highlights the importance of accurate estimates of the basic sector. The Keynesian approach is but one way to think of economic linkages and multipliers; alternatives are discussed in detail in Chapters 15 and 16.

**INTERNAL MARKETS: CENTRAL PLACE THEORY**

Export base theory argues that the nonbasic sector is totally dependent on the export base, but in reality, such is not the case. The provision of local goods and services by local firms greatly affects the well-being of community residents. Two dimensions of supplying these goods and services are of particular interest for community economic analysis. First is the interdependence among merchants across different communities. As we saw in the demand-maximizing approach to firm location in Chapter 3, firms are dependent on where their competitors locate. Second is the firm’s interpretation of market demand and the socioeconomic characteristics that dictate market demand and how that interpretation influences decisions to provide particular goods and services.

Central place theory (CPT) is one conceptual framework that addresses these two facets of economic activity. CPT does not provide all the details needed to determine the feasibility of a particular investment in a particular market, but the theory does help to collect the necessary details needed for competition analysis. Central place theory offers insight into why specific goods and services are or are not present in a particular community. This theory specifically recognizes that no community’s trade, or nonbasic, sector can be viewed in isolation. It is important to remember that central-place theory essentially argues there is a hierarchy of communities based on the functions (retail and service) that are provided in the community. In a sense, central-place theory attempts to place local goods and services margins of our circular flow model onto an economic plane.

Central place theory takes our demand maximization problem of Chapter 3 to the next step by defining a spatial system of markets or communities (Mulligan 1984; Par 2002; Potter 1982). In Chapter 3 we outlined how an individual firm can increase sales by using a more effective price that generates more demand per person or through an increase in the geographic area included in the firm’s market. A circular market surrounds each firm (Fig. 3.5). Transportation costs, consumers’ willingness to pay the effective price, and competition from other locales determine the limits of each market. This system of circular markets leaves some areas unserved. Perfect competition allows competitors to enter the market, serve presently unserved areas, and compete away the excess profits of current firms. Thus, entry of competitors creates a series of regular, hexagonal market areas (Fig. 4.3).

Two critical behavioral assumptions of central place theory are that businesses will attempt to maximize the area served, and that consumers will attempt to minimize the distance traveled. What we see in Figure 4.3 is how firms will allocate themselves on a homogeneous economic plane in a way that maximizes profits by maximizing demand. We also begin to see the economic foundation for a system of places on the economic plane. As we introduced varying cost structures of different types of firms below, we start to develop a hierarchy of urban places. Two concepts of particular importance to this hierarchy that play a fundamental role in community economics are the range and the demand threshold of a good or service.

**Range of a Good or Service**

The range of a good or service is the maximum distance people will travel to purchase that good or service at a particular location (Berry and Garrison 1958b; Olsson 1966; Farr and Denike 1970). The range is the outer limit of the geographic market for a good or service from a particular location. In the Lösch demand cone (Fig. 3.5), the radius of the demand cone is the range of the good or service. The circumference of the demand cone defines a trade area, which is the geographically delineated area containing potential customers to purchase goods and services offered for sale by a particular firm (Davies 1977; Huff 1964).

The distance determining the geographic limit of a market is measured in terms of physical separation and travel costs, including time. Other determinants of the outer geographic limit are the ease of access to competitive markets, transportation facilities, and technology (Shepard and Thomas 1980). Ease of access to competitive markets reduces the geographic limits of the market area for a community. Physical features such as mountains and rivers influence access. Transportation facilities and technology are important because better facilities and faster movement permit greater spatial movement by consumers.

Individual characteristics influence the distance people are willing to travel (Shepard and Thomas 1980). Younger, more educated, higher-income people are likely to travel farther and more frequently than those with contrasting characteristics. The distribution of income affects the number of people able to pay transportation costs of greater movement. Movement imagery involves the consumer’s perceived options about movement from one place to another in the quest of a desired good or service. The mode of travel, travel time, the cost of overcoming distance, and communication flows affect movement imagery. Behavioral space is that part of the total central place system which the individual perceives as a potential source for satisfying his or her demand for goods or services. This behavior space is influenced by previous shopping experiences in the various central places and by sources of information, such as advertising. For any given consumer, the behavior space can include several communities.
market these two firms will require different demand structures to yield a competitive equilibrium. The demand thresholds, or the population required to support the two separate businesses, will be different, with Firm 2 requiring the larger market. If consumers are evenly distributed across our economic plane, the larger market required for Firm 2 will be a larger geographic area, or larger range.

Consider, for example, two establishments: a tavern and a high-end furniture store. The tavern has only a handful of costs associated with its operation: the rent of the building or space, the cost of goods sold, labor, and miscellaneous costs, such as utility and insurance costs. A high-end furniture store, on the other hand, has a much higher cost of goods sold and requires a larger building, higher level of labor-related services, such as delivery and design services, and advertising. In Figure 4.4, the tavern may have a cost structure similar to AC₁, while the furniture store may have a cost structure similar to AC₃. Clearly, the furniture store will require a much larger market area or demand structure (AR₃) to support its operations than will a tavern (AR₁). In other words, the range and threshold of the furniture store will be significantly greater than the tavern. Given different ranges and thresholds for different types of firms, we begin to see a system of overlapping markets. These overlapping markets form the basis of a hierarchy of central places.

Central Place Hierarchy

Recognizing that central places and their tributary areas cover the entire market surface of our economic plane, the elements of a system of central places emerge (Berry and Garrison 1958a, 1958b). The theory provides a framework that helps explain and predict where heterogeneous firms will locate and cluster together. The classic presentation of Christaller’s (1933) system of central places shows places (cities) arranged in tiers with each member of the tier, except the highest tier, subordinate to at least one other central place on a higher tier (e.g., hierarchy) (Fig. 4.5). The lower levels of the hierarchy are made up of more central places serving smaller tributary areas. Moving up the hierarchy, there are fewer central places, but each serves an increasingly larger geographic area and population than the communities (central places) at lower levels. In the real world there are numerous small hamlets or villages (lower-level places) but only a small handful of large cities (higher-level places).

A mutual dependence exists between the central place and its tributary or complementary area. The central place provides higher-order goods and services and the tributary area provides a market. The tributary area contains lower-order central places for all but the lowest-order center, which contains just farms or scattered site residences. For example, the market area of a hamlet or village is relatively small,

![Figure 4.3. Formation of hexagonal market areas.](image)

**Demand Threshold**

Demand threshold is the minimum market required to support a particular good or service and still yield an abnormal profit for the merchant (Berry and Garrison 1958a, 1958b; Deller and Harris 1953; Henderson, Kelly, and Taylor 2000; Olson 1966; Parr and Denike 1970; Shookwiler and Harris 1996; Wensley and Stabler 1998). The concept of demand threshold, based on the internal economies of the firm and the characteristics of consumer demand, is defined where average cost is just equal to or tangent to average revenue (Fig. 3.6). Because of this, demand thresholds are not absolute; they vary with the type of good or service. Demand thresholds usually are measured in terms of population, rather than quantity sold, by assuming consumers are homogeneous in their buying power (income) and tastes.

The internal cost economies of the firm determine the thresholds that will yield a competitive equilibrium (Parr and Denike 1970; Olson 1966). Suppose we have two firms that offer different goods and services, hence they have two separate cost structures (AC₁ and AC₃ in Fig. 4.4). Clearly, in a competitive market, the lower-cost firm will be able to capture a larger share of the market.
Figure 4.5. Idealized central place hierarchy.

but the market area for a large city is much larger and includes the small hamlets.

The tributary areas for different central places in a central place system become a collage of overlapping boundaries (Fig. 4.5). The larger cities provide more specialized activities to a larger tributary area. At higher levels of the hierarchy, the number of central functions increases and they become increasingly specialized (Fig. 4.6). These cities also provide the same general services found in the smaller cities. The smaller places offer more generalized goods and services to geographically restricted trade areas. Each level of the central place hierarchy has its own system of hexagonal market areas. As a result of the demand-maximizing problem from Chapter 3 (in which consumers react to an effective price that includes travel costs, and firms maximize profits) that is the center of many market areas will occur at the same time regardless of the hierarchical level. This offering of several goods and services from one location prevents a totally random settlement pattern and increases the efficiency of providing each good or service.

Suppose that in Figure 4.5 we have five different types of firms that vary in terms of cost structure, with H-type firms having the lowest cost structure, such as a tavern, and G-type firms having the highest cost structure, such as a luxury car dealership. As we have seen, the population required to support H-type businesses will be much smaller than that for G-type businesses. Accordingly, on our homogeneous economic plane, we would expect to find a large number of places that have H-type businesses but only a few that have G-type businesses. Firms will enter the market and locate on the economic plane in such a way that all consumers are captured and profits are maximized. In this example, the spatial economy may be large enough to support only one luxury car dealership, which will locate in the center of the economic plane and offer services to the entire plane. If the same time, the spatial economy can support numerous taverns evenly distributed across the economic plane. If we want to, we can relabel business types with place types. We can call G-type places cities and H-type places small hamlets or villages.

The number of different economic (central) functions available differentiates places. The number of central functions performed depends directly on the cost structure of businesses and the population of the spatial economy. Since costs vary among types of goods or services (i.e., central functions), some goods or services are available only in the higher-level places, while others are available in even the smallest places. There is a pattern of similarity between market size within a given hierarchical level and the differences between hierarchical levels (Foust and de Souza 1978).

The regularity and flexibility of how places come to be described as central is outlined in Figure 4.6. Every hamlet has a gasoline service station and a tavern/resting place. Every minimum convenience place has all the central functions found in a hamlet plus a grocery, drug store, hardware store, and bank. In addition, a minimum convenience place should have any two or more of these central functions: garage/auto and implement dealer, a variety store, a meat/fruit/market, and a general merchandise store. Different authors use different labels for the various levels of the central place hierarchy, but the labels, per se, are less important than the recognition that different levels of the hierarchy provide a different mix of goods and services. From the point of view of business, one reason for the hierarchy is conditions of entry. With free entry and the minimum market required (threshold) to support the business available, the business will exist. It cannot exist without the minimum required market. Both minimum market and cost structure of the firm are linked to the number of customers or volume of sales in the community and available to the business (i.e., not already served by competitors).

A key assumption of central place theory (as summarized in Fig. 4.5) is an even distribution of homogeneous people across the economic plane. Clearly, this assumption does not hold in the real world; locating the assumption greatly complicates our notion of central places, but it does not destroy it. People cluster together into cities and villages, fundamentally altering the shape of our functional demand cones. Likewise, external economies of size vary with each community, but each one is not independent of the economic characteristics. This community influence is often expressed as external economies or economies of agglomeration that shift the firm's average cost curve. By allowing for something other than a homogeneous economic plane, the rather sterile view of central places
becomes much more reflective of the real world. The ideas of multipurpose shopping trips, business clustering such as shopping malls, and market segmentation such as entertainment districts as opposed to office buildings begin to make sense. The wealth of these concepts is dissolved in a policy context below.

One immediate outcome of lifting the homogenous economic plane assumption is the use of population as a proxy for differentiating cities within the hierarchy. The logic for substituting population for an actual tabulation of central functions is that the market required to achieve economies of size and make certain goods and services available is linked to population. Returning to Figure 4.3, we now cast think of places in terms of population rather than number of businesses.

Consumers play a part in creating the hierarchy of central places as well. First, consumers wish to minimize the distance traveled to purchase any good or service. Therefore, frequently purchased items should be available nearby; less frequently purchased items need not be located nearby. Consumers' desires to minimize total travel means that they go to higher-order centers only for the goods and services not available in lower-order centers. Even people who live in higher-order centers will purchase lower-order goods and services in their neighborhood. Now make only single-purpose shopping trips to the higher-order centers.

Second, the type of shopping also influences the hierarchy of goods and services. Goods and services which are everyday convenience items (e.g., milk, bread) will be available in smaller central places, but if an item is subject to comparative shopping (furniture, automobiles), it is likely to be available only in larger centers.

Third, the type of transportation system available affects the hierarchy through the frequency of shopping and spatial dimension of the market area. Important elements of the transportation system include whether individual (car) or mass transit is the source of transportation. Consumers dependent on public transportation may find geographical as well as time-of-day limits placed on shopping. Both factors limit the market available to businesses in a central place.

Shifts in the Hierarchy of Central Places

One of the powers of central place theory is its ability to help the community practitioner think through how changes of central places will filter through the retail and service markets. As we have seen repeatedly, communities do not function in isolation from their surrounding communities. Any change in the economics of one community will affect not only its own place in the hierarchy, but all communities in the hierarchy.

The community practitioner needs to understand how shifts to one community can affect communities in the system. For example, what happens to the system of central places if the population of one place significantly increases? Consider a case where a small hamlet has a large manufacturing firm locate which in turn draws a significant increase in population. Clearly, the hierarchy of central places will shift, but how? In the simplest sense, the hamlet's place in the hierarchical chain will move upward. The hamlet will attract more businesses because the higher population will satisfy the demand thresholds of more types of businesses. But equally important is the impact that change will have on other communities in the hierarchy. In essence, the spatial competition of firms and markets has changed. Neighboring hamlets will find a sharp increase in competition.

Another change or shift might be a decrease in transportation costs. Clearly, as transportation costs decline the effective price facing the consumer will decline and the consumer will be willing to travel greater distances to make purchases. Not only will a decrease in transportation costs expand the spatial market of one place, but of all places. Investments in transportation infrastructure is a two-way street: Not only do local businesses see an increase in the spatial size of their market, but also local customers can now more easily go to other markets.

We can also use central place theory to help think through the impact of changing household, demographic, and retail and service markets. With time becoming an increasingly valuable commodity, consumers will look to maximize the effectiveness of their time spent shopping. Consumers will look for places or communities where multiple purchases can be made. As we have seen in our discussion of central place theory, higher-order places will offer a greater variety of goods and services. With the increased pressure on the value of time, these higher-order places will be more attractive. The advent of shopping malls and big box retailers in higher-order places feeds into economic externalities of size and agglomeration, giving higher-order places an even greater comparative advantage over lower-order places. The prediction of the theory would be that the system of central places will be more attractive.

Using the concepts of the range and the threshold of a good or service, we can think through changes to the local community and the impact of those changes on the surrounding communities. Perhaps most important is the improved understanding of how local communities are interconnected in a predictable manner.

Limitations of Central Place Theory

While central place theory offers considerable insight into the spatial allocation of consumer functions, its limitations must be recognized (Parr 1973; Parr and Denske 1970; Shepard and Thomas 1980; Turner and Cole 1980).

In its earliest form, central place theory assumed the physical dimension of space (geographic space) was the dominant element. The consumer's perception of economic space was not considered. In this form, central place theory was limited because in the real world, physical distance is irrelevant; consumers and producers use travel time and its associated cost. In other words, the real world is driven by economic space, not by geographic space.

Originally, Christaller assumed that the hexagonal markets of central places were on a homogeneous plane with a uniform population. Later, Löschian adaptations erode economic distortions that affect the size-and shape of these hexagonal markets. The transportation network (roads, railroads) was shown to have dramatic influence on the size and shape of markets. Also, an unequal distribution of natural resources and physical barriers, such as rivers and mountains, affects the shape of the market area. Economies of size and different population densities alter market area size but not shape. This means that new transportation technology which reduces travel time or cost or both and which alters consumers' perceptions of access to a given place will alter the market size because physical distance is unchanged.

Another limitation of central place theory is that it does not consider the qualitative diversity among alternative locations (central places). Qualitative dimensions include product selection, parking availability, store hours, clerks' attitudes, community quality of life, and amenities in general. In addition, the idea of a tourist visiting a community's downtown is a completely alien concept. As we will see below, the market for tourism can be a significant part of a community's strategy for development of its internal markets.

The theory implicitly assumes that an equilibrium exists between demand and supply and that this equilibrium will be maintained. Thus, it provides very little insight into the adjustment process that takes place after some economic disturbance. Since the spatial distribution of consumers is given, the theory doesn't explain why changes in population would occur. Fluctuating residential settlement patterns make it difficult to determine if a market hierarchy will exist in a given market area. Central place theory is limited in explaining the causal element behind agglomeration; the theoretical construct is silent on the underlying reasons why change might occur. Central place theory fails to explain the evolution of places among different levels of the hierarchy (Parr 1981). The theory is potentially dynamic; it uses comparative statics to explain conditions not changes.

A final limitation of central place-theory is the assumption of single-purpose shopping trips. The exclusion of multipurpose shopping trips distorts results from the analysis (Parr 1973). If the consumer makes multipurpose trips, the total cost (purchase price plus transportation) for any item is lowered by sharing the transportation costs between more than one good or service and thus increases the real range of all goods and services. This phenomenon is particularly significant for stores offering lower-order goods and services in higher-order central places; such stores can acquire excess profits because their market contains those individuals who would be good or service had to bear the full cost of consumer travel. This, of course, assumes the bidding for labor or land in the central place does not shift the average cost curve. The spreading of travel costs over several goods and services purchased on a multipurpose shopping trip reduces the price advantage of firms locating near the consumer (i.e., reduces the effective market where the local firm has some monopoly price control). The assumption that consumers have complete knowledge of their spatial options to purchase a given good or service eliminates the consumer risk reduction strategy of bypassing a nearest intermediate center, which may or may not have the item, and making the purchases in a higher center (Shepard and Thomas 1980).
commercial strips. Many downtowns face high vacancy rates, a poor mix of retail tenants, and multiple property owners. Periphery shopping areas are characterized by newer investments in private and public infrastructure and major land use changes. Communities and individual merchants typically lack the market research support available to the big retailers and shopping center developers. The citizens and business interests of the community need to come together and study and reflect both on market conditions and preferences of the community in determining the type of retail/service they desire.

Understanding market conditions is the first part of any analysis (Chapter 14). This includes analyzing current building uses, business mix, trade area size, economic (including competition) and consumer data, consumer attitudes, and business operator needs. When analyzing internal markets, always remember to analyze the competition from larger and/or nearby areas. This will provide the necessary foundation for more in-depth analysis of different business sectors. Market opportunities are identified by specific business sectors, including retail, service businesses, restaurants, entertainment opportunities, residential units, office space, and lodging facilities.

How Trade Areas Differ

As central place theory suggests, different business types will have different trade areas. That is, people will travel from greater distances to purchase certain goods and services than others. While each individual store may have its own unique trade area, these areas can often be generalized into two different types: convenience-shopping trade areas and comparison-shopping trade areas. Local convenience trade areas are based on the ease of access to these types of products. That is, people will obtain these products (e.g., gasoline, groceries) based on travel distances or travel time. Conversely, comparison-shopping trade areas are based on price, selection, quality, and style. People are more likely to compare these types of goods (e.g., appliances, furniture) and to travel longer distances for their purchases.

In addition to different types of shopping goods, there are also different types, or market segments, of customers frequented a downtown. Three common market segments are local residents, daytime employees, and tourists. Local residents live within the trade area. As they live year-round, they provide the majority of spending potential for most downtowns. Daytime employees may live in the trade area but may also commute from other outside areas. While these employees are in the downtown, however, they provide the potential to stay and make purchases. Furthermore, depending on the community, tourists can provide a large amount of spending potential. While they are not permanent customers, tourists make purchases while they visit the area. Indeed, for many high-amenity areas, the tourist can dominate the internal market.

Identifying High Potential Internal Market Sectors

Communities can influence the form and timing of retail and service sector investment decisions by drawing attention to local market characteristics that might otherwise be overlooked. Economic development professionals can assist by identifying high potential sectors, compiling information of interest to prospects, and marketing to those prospects. The key is to demonstrate that the community is a profitable place to do business (Shaffer and Ryan 1997). The key to any market analysis is that the location and business type must find a competitively advantageous place (i.e., a profitable place to do business). Porter’s (1995) fundamental argument is that economic activity “will take root and grow when it enjoys a competitive advantage and a niche that is hard to replicate elsewhere.” Porter’s analysis reinforces the importance of focusing on unique community characteristics.

To identify appropriate retail and service sectors in a community, it is often useful to first analyze retail and service deficits or opportunities. Analysis of the local context is critical to sustainable retail and service sector development (Ryan and Campbell 1996). We need to ask questions like, Why did this concept work in community X? Are the same factors present in our community? Often we look to other places for ideas on how we can revitalize retail and service sector activity in our community. As a result, we may overlook the unique characteristics of our community that can lead to successful and appropriate retail and service sector development.

If there appears to be demand in certain retail and service sectors, competition in and around the trade area must be carefully evaluated so that an oversupply of a certain type of business is avoided. Do not try to attract businesses to your community if demand for their products is too low or competition is too fierce. Local or regional businesses, particularly those that have branch locations, are often excellent prospects for expansion. They typically have a good knowledge of the market area, and if they already have multiple locations, have demonstrated an interest in expansion. They are often interested in expanding as a way to improve their penetration of the market. Retail and service sector trade shows and conferences or industry newsletters offer a direct means of contacting potential businesses.

Information of Interest to Retail and Service Sector Prospects

Retailers are very interested in knowing the size of the market for their products or services, and how effectively they can penetrate that market. They want to minimize their risk by selecting sites that offer the greatest sales potential. Communities can provide data to prospective retail/service businesses. Data can help potential retailers:

- Demographic data, such as population, age, income, and ethnicity.
- Lifestyle data profiling buying behavior of local residents.
- Local construction trends in housing, commercial, and industrial space.
- Local employment trends.
- Transportation data, including traffic volume and parking.
- Mix of existing retail/service, entertainment, and services in the area.
- Mix of residential, lodging, office, and industrial space in the area.
- Local and regional competition, including location and size.

When targeting retail and service sector prospects, remember that not all businesses have the same requirements. As we saw in central place theory, a grocery store typically requires different market characteristics than a hotel. Communities should customize information to fit the needs of the prospect.

To attract internal market firms, a community must first make its business district visibly attractive, attractive, convenient, and safe. This is often more difficult for non-shopping center locations, including downtowns, because they do not operate under central management. It is important to get local merchants organized early on to address issues like hours of operation or sidewalk cleanliness or safety. It is also important to highlight what the community is doing to increase activity in its business district. This might include government incentives, including TIF districts, façade improvements, tax abatements, and lending programs. If available, a master plan for the community should be provided that explains how new development, parking, traffic, security, beautification and cleanliness will be handled.

Retail and Service Clustering

Clustering builds on the agglomeration effects within and across retail and service sector markets. In our context here, clustering is a more narrowly defined concept than that introduced in Chapter 3. Much of the current research on central place theory is focused on clustering of retail and service sector business and spatial competition (Henderson, Kelly, and Taylor 2000; Schankwiler and Harris 1996; Wansley and Studier 1998). Retail and service sector clustering is an important but often overlooked feature of business recruitment strategies. Clustering is the grouping together of a mix of businesses that enable individual businesses to benefit from each other’s sales and customers. Clustering is a technique long used by shopping centers and retail district developers.

Clustering provides consumers with a critical mass of businesses in one location and creates retail and service sector synergy. Clustering can:

- provide consumers with a broad selection and variety at a single convenient location.
- enable consumers to make purchases at more than one business and satisfy a number of shopping needs in one trip.
- allow a business district to function as a single economic unit instead of a series of unrelated destination businesses.
- increase spending since the appropriate mix of businesses will offer more goods and services that appeal to targeted shoppers.
- increase impulse buying among clustered stores that offer complementary goods.

For business clustering to be successful, an appropriate business mix is essential. Individual businesses must be able to effectively serve the same or overlapping segments of the market. Clusters also must be physically located so that they are compact and are not interrupted by incompatible space uses. The cluster must encourage the customer to shop the entire cluster.

When developing a cluster policy, it is important to understand that there are three basic types of clusters:

- Compatible clusters are groups of businesses that share a particular market segment but offer unrelated goods and services. Outlet malls are an
example since their tenants share a market segment that enjoys looking for bargains. Most business districts are classified as compatible clusters.

- **Complementary clusters** are groups of businesses that share customers and market segments but offer complementary goods and services. An office supply store, copy center, and office furniture store together could form a complementary cluster (business services). Retailers must offer goods and services of a similar style, quality, and price range. It is interesting to note that department stores are typically organized this way.

- **Comparison clusters** are groups of businesses that carry the same or similar goods and often appeal to the same markets. In some larger regional malls, a clustering of shoe stores can be found. Consumers are able to shop the various lines and compare goods before purchasing them. This also is observed within many department stores.

### Developing a Clustering Strategy

Clustering in malls and shopping centers is relatively new because they are site and/or merchandising plans in place from day one and, more importantly, have one owner. They have the flexibility to move or resize their tenant's space and replace tenants that no longer fit into the overall mix. Traditional commercial centers, such as downtowns, have multiple property owners, some of which do not live in the community. Business leaders need to overcome this obstacle and show property owners the benefits of clustering, namely the maximizing of real estate values, which occurs in successful clusters.

Hyett-Palma (1989) recommends a four-step clustering strategy for business districts that do not have the centralized control of a shopping center. First, one needs to analyze the market served by the business district to determine the targeted markets and appropriate mix of businesses for the district. This should address the trade area, target market purchasing characteristics, competition, character of existing businesses, image of the center, projection of realistic sales capture potential, and appropriate mix of businesses. Customer surveys can be used to reveal undeserved retail and service segments within the area.

Second, one needs to prepare business-clustering maps for the business district. This includes maps that display (a) existing businesses and available commercial space, (b) what types of clusters and their locations might be appropriate for the business district, and (c) the specific types of businesses as well as the optimal placement within the center given available space.

Some combinations of retail and service sectors do not work well together (Ray 1996). For example, some apparel retailers are not good together with grocery stores because shopping for clothing and food is seldom done on the same trip. When filling space, it is important to know what types of stores are complementary.

Third, the community needs to gain control of the building space within the business district if possible. This could be done by centralized retail and service sector management by a group of property owners and businesses that provides a coordinated set of activities, including implementation of a leasing plan; having the business district organization obtain the right of first refusal to approve or disapprove new tenants; or obtaining voluntary cooperation by showing the owners that they can benefit from a viable mix of businesses.

Fourth, the community needs to institute an aggressive management mechanism for the business district. The lead organization must have the support of businesses, property owners, and local government officials.

### SUMMARY AND POLICY IMPLICATIONS

Community economic analysis rests on a conception of markets. Markets can take many alternative forms. In particular, two market conceptualizations dominate the thinking of how community economic activity is linked: the market for goods and services and the market for factors of production. Households demand goods and services in the former and supply factor inputs in the latter. Firms, on the other hand, supply goods and services and demand factor inputs.

The focus in this chapter is on the market for goods and services. Two specific theories relate to this market. Export base theory emphasizes the importance of producing goods for external consumers and the feedback on internal markets. Central place theory, on the other hand, focuses on internal markets and on the interrelationship among communities that make up a hierarchical system of places.

Historically, most communities have pursued a policy of maintaining and expanding businesses that export. In essence, these activities act to attract new dollars into the community from outside. The assumption of these economic development approaches is that the businesses that do not export (i.e., the nonexporters) will grow at a rate determined by the amount of exporting that takes place. In other words, internal markets are completely driven by the external demand for exported goods and services. The spending and linkages that occur between the export and nonexport sectors do not need to appear locally. In essence, from central place theory, the amount of local consumption will be driven by a community's place within the economic hierarchy.

The significance of central place theory to community economic analysis is its recognition that the community is part of a larger urban hierarchy. No community, especially a smaller community, can provide all the goods and services necessary and desired. Residents in smaller communities and their surrounding tributary areas necessarily relate to larger communities for many goods and services. The community economist must recognize the relationship between the range of a good or service and the demand threshold for a good or service. The range of a good or service indicates the geographic limits of the market area for that central function. The threshold indicates if there is sufficient demand within that market area to justify offering that particular central function from a certain central place.

Those who perceive the market for goods and services as being dominant in community economic development maintain that the lack of long-run economic development is not the result of inadequate productive capacity. They point out that communities are often faced with significant unemployment, unused capital, and population migration or commuting out of the community. Rather, those that emphasize the market for goods and services contend that the lack of development in a community results from inadequate consumption demand. What communities need is to increase either the internal or external demand for goods and services produced by the community, or change the types of goods and services produced in the community to better match existing demand. This is essentially a structural question.

As communities make decisions about their economic future, it is important that policy-makers understand the practical importance of the two theories discussed in this chapter. Public and private decisions that affect business activity will inevitably create change in markets. The most effective economic development decisions are crafted within a broader theoretical understanding of how markets react. A market approach argues that policies and decisions that solely treat factor markets tend to be incomplete. The real problem is that the community needs to produce goods and services for which there is sufficient demand. These demands can exist as both internal and external to the community. Communities wishing to grow are best served by focusing production on goods and services where demand is growing.

### STUDY QUESTIONS

1. **The circular flow of a community's economy represents a flow of what economic characteristic?**

2. **How do households and firms play different roles in the supply and demand structure of alternative markets?**

3. **What is the significance of the export sector and how does it affect community economic development?**

4. **How is an export transaction defined for community economic analysis? What forms does it take?**

5. **What are some of the assumptions made when using export base theory?**

6. **Does the export base theory indicate that there is sufficient demand within the market area to justify offering a particular central function from a central place?**

7. **Why is it necessary to assume a constant structure between export and nonexport sectors?**

8. **Why is the determination of the size of the export sector so important? What are the implications of having a homogeneous export sector?**

9. **What are some of the forces that increase the export sector or even allow the creation of new export sectors?**

10. **Some central place terminology includes central places, range of a good/service, and demand threshold. What do these terms mean and how are they linked to community economic analysis?**

11. **What are the implications of central place hierarchies to community economic analysis, especially the provision of goods and services? What behavioral assumptions for firms and households are required?**

12. **Why might demand thresholds vary among places, over time and among different goods/services?**

13. **What is a trade area?
Section II
Community Factor M

The building blocks of community economic development exist as resource endowments. This section focuses on community resource markets that are important in understanding how regional and community output is produced. Often referred to as primary factor inputs, this section includes separate discussions related to the standard tangible factors of land, labor, and capital, but also develops an appreciation for the more latent factors that are important to community economic development. These latent factors include technology and management, amenities, and publicly provided goods and services. Another way of looking at these factors is to think of the bottom half of the circular flow model.

Factor input markets are critical aspects of community economics for a variety of reasons. In addition to providing the basis for producing goods and services, their employment represents a critical aspect associated with how households generate income, how society inserts itself within the market framework, and how local decisions can control the extent of community quality-of-life attributes.

The section begins with three chapters that outline the traditional primary factors of production: land, labor, and capital. Indeed, the employment of these three inputs provides households with income to spend on consumption goods and services. Our basis of land value focuses on alternative conceptions of land rent. The spatial components of land rent play dual roles of explaining income generation opportunities for landowners while simultaneously explaining the spatial array of alternative land uses with respect to markets. This latter element is critical for developing a basis to explain contemporary suburbanization and exurbanization pressures within which land is a central issue.

Community labor market decisions made by the most obvious element. Our discussion of is supply and demand components with a particular focus on key development issues of community labor markets. These include labor mobility, unemployment, exploitation, and discrimination.

Our initial discussion of capital markets focuses on private financial capital. In distinguishing stock assets from flows, we lay the framework for discussing capital from both a debt and an equity perspective. We conclude with a discussion of several key capital market failures that focus attention on development issues such as mobility, risk and uncertainty, and regulatory influences.

Although important in explaining community economic activity, the traditional primary factors of production are insufficient when understanding the unique aspects of development. This is particularly true in smaller communities. The more latent and somewhat less tangible factors associated with technology, amenities, and publicly provided goods and services are critical components of the modern vibrant and increasingly affluent economic plateau that we've achieved at the onset of the 21st century. In order to capture this set of critical latent factors, we've added the last three chapters to the section.