

2

Economic Methods of Analysis

“Two and two the mathematician continues to make four, in spite of the whine of the amateur for three, or the cry of the critic for five.”

James McNeill Whistler

Most of the underlying theories of facility location and land use models are basically economic concepts, and many of their input/output variables are economic measures. To understand these relationships better, a general knowledge of economic concepts and methodology is helpful. We recognize that theories have been offered by economists to explain the growth and distribution of industrial activities in an area. It is insightful to summarize their experiences—particularly the theories used in regional and interregional economics. This includes such concepts as economic-base theory (or export service theory) of gravitational interaction and theory of interregional flow. Through such a review, one sharpens the focus on the validity and limitations of these analysis methodologies.

We will also outline the basic techniques for evaluating the impact of a proposed policy on transportation systems, utility systems, and zoning codes. When an evaluation measure is often phrased in terms such as cost, benefit, equity, and efficiency, a clear understanding of these terms is necessary. Conversely, when indicators such as opportunity and quality of life are output from the model, they are much more meaningful if one can relate them to the economic theories of cost/benefit and equity/efficiency. Such an understanding would help the inquiring mind to understand the assumptions based upon which the measures are derived. Finally, for the model builder, the review of economic methods would help them configure better models and submodels.

I. ECONOMIC CONSTRUCTS FOR ACTIVITY ALLOCATION AND FORECASTING

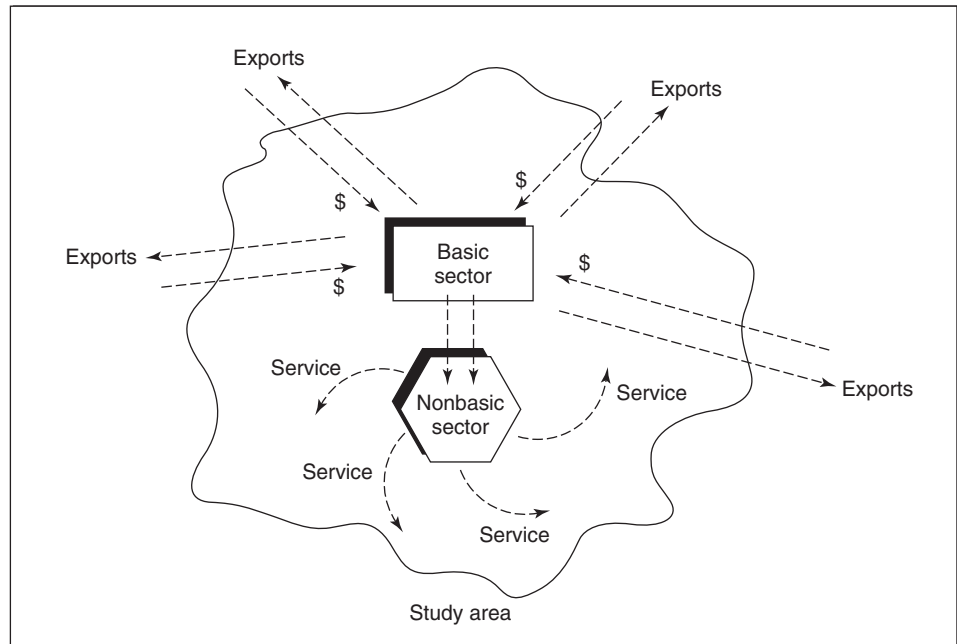
Econometricians have been forecasting economic activities such as population and employment for a long time. Two types of forecasting methodologies can be broadly classified—forecasting on the basis of **cross-sectional data** versus that

based on **time-series data**. Using cross-sectional data, models are calibrated on the current spatial distribution of activities, thus examining a “snapshot” of the population/employment distribution on the map. A time-series approach, on the other hand, would utilize not only the current pattern, but also previous patterns, which allows an observation over two or more time periods. The former is a static way of forecasting, while the latter is more dynamic. In other words, the former assumes the general activity distribution pattern will prevail over time, whereas the latter recognizes explicitly that changes over time are an integral part of the development. Aside from their important role in the development literature, the three economic concepts—economic-base theory, location theory, and input-output models—are selected for further discussion because the first two illustrate cross-sectional forecasting methodology, while the last one illustrates time-series forecasting.

A. Economic-Base Theory

The term **economic base** has many different usages and meanings so that it is necessary to clarify the definition for use here. In general, the term economic base has been applied to activities thought of as being major, fundamental, or of considerable importance in the economic structure of an area. The economic base of a community consists of those economic activities that are vital to the continued functioning and existence of that community. An economic-base study is an attempt to determine those economic activities devoted to the export of goods and services beyond the study area’s borders. This activity is thought of as being the primary reason for the earning ability and economic growth of the community. Because these basic industries sell their products and services outside of the area, nonbasic or service industries can be supported within the community’s boundaries. For example, barbers, dry cleaners, shoe repairers, grocery clerks, bakers, and movie operators serve others in the area who are engaged in the principal activities of the community, which may be mining, manufacturing, trade, or some other industry. These service industries have as their main function the provision of goods and services for persons living in the community.

This distinction of basic and nonbasic sectors of economic activity in an area is illustrated in Figure 2.1. Note that the income of the nonbasic sector is dependent upon the income of the basic sector so that it seems that the service industries only exist to serve basic workers and other service workers. Hence, fluctuations in income or employment in the basic sector will ultimately affect income and employment in the nonbasic sector. Since the nonbasic sector activities depend upon the basic sector, changes in the basic sector will have a net effect on the entire study area economy when some multiplier is applied to the economic-base method of analysis. The economic-base multiplier attempts to predict the change that will occur in the study area economy given a forecast of changes in certain basic activities. A significant part of the analysis involves the construction of these impact multipliers. They are numerical constants intended to impose the effects of changes in the demand for an area’s goods and services upon the volume of employment or income in that region. For example, a government contract for a defense item increases employment in a firm by 2000 jobs. Indirectly both contract and job increases might generate still more work opportunities and produce a total increase in local employment two or more times a multiple of the original 2000.

Figure 2.1 CONCEPT OF BASIC VERSUS NONBASIC ACTIVITIES

SOURCE: Adapted from Newman (1972). Reprinted with permission.

Example

Using employment as the unit of measure, classify the employment of all industries in the study area as basic or nonbasic. Establish the Normal Ratio, the relationship between basic and nonbasic employment that usually exists:

$$\text{Normal Ratio} = \frac{\text{Nonbasic Employment}}{\text{Basic Employment}} \quad (\text{Assume a 2:1 normal ratio, for example.})$$

Total Employment = Nonbasic Employment + Basic Employment. Assuming the total study area employment to be 90,000, then nonbasic employment is now 60,000 and basic employment is 30,000.

$$\text{Multiplier} = \frac{\text{Total Employment}}{\text{Basic Employment}} = \frac{60,000 + 30,000}{30,000} = 3$$

If basic employment is forecast to increase by 15,000, the total increase in nonbasic employment would be $3 \times 15,000 = 45,000$. Then the total employment for the forecast year becomes $15,000 + 45,000 + 90,000 = 150,000$. Since the normal ratio of 2:1 still holds, nonbasic employment is 100,000 and basic employment is 50,000. ■

Thus, economic-base theory is to describe the development of economic activities in a typical area or region. The development of economic activities in a specific area can be explained in terms of the following four stages:

- Step 1:** Calculate the total population and employment and the amount of constituent basic and nonbasic (service) employment;
- Step 2:** Estimate the proportion of basic employment to population and that of population to service employment;
- Step 3:** Estimate the future trend in the basic employment; and
- Step 4:** Calculate the total employment and total future population on the basis of the future trend in basic employment.

In other words, basic employment has to be determined exogenously, then based on the multipliers such as labor force participation rate and population-serving ratio, which are the two proportions mentioned in Step 2, future employment and population in the region are estimated. Aside from the example above, another numerical example of the economic-base concept was given in Chapter 1 in Table 1.1.

The validity of future estimates of employment (or any other variables) depends upon the relative stability of the nonbasic-to-basic ratio developed. However, the economic-base method still has many problems to be solved. Some of these are:

1. Determining which activities are basic and nonbasic;
2. Choosing which units of measurement best represent the economy; and
3. Establishing the geographic area boundaries for which the base study is to be made.

In addition to these conceptual problems, other criticisms of the economic-base method have been registered. As the size of the study increases, the ratio of nonbasic to basic employees increases with a resultant increase in the multiplier. As a consequence, large areas have very large multipliers which do not truly reflect total economic change due to changes in the basic sector. It becomes apparent that the economic-base multiplier method is most applicable to relatively small areas and towns. Some critics challenge the premise that basic activities are more important than service activities because of the important contributions of such factors as the transportation system, communications network, and other systems serving the community. This criticism is important because planners use the basic-nonbasic distinction to emphasize which industries should be built up to improve the community's economy and to improve the balance of payments. Industries that produce goods which are presently imported would be neglected under this premise. More technical treatment of the subject will be found in Chapter 3.

B. Location Theory

Location theory, a study of the effects of space on the organization of economic activities, is a body of knowledge about the location of different activities or the rationing of different resources so as to achieve desirable spatial interaction. It has its genesis from early studies of the relative locations of plants and industry, in which the availability of raw material and the accessibility to consumer markets are of primary importance. According to the spatial price theory, transportation cost is the price for rationing resources and economic activities. For example,

manufacturing plants and industries find the most convenient locations at close proximity to the input resources (both labor and raw materials) or consumer markets in order to minimize transportation costs. Another good example is a family's choice of housing location, in which a tradeoff is made between the transportation costs and other expenditures and values. If a heavy weight is placed on freedom from the noise and rush of the central city, the family locate at a distance away from the city and pay the transportation cost. In their decision, the utility of a serene environment is much higher than the utility of being close to jobs and other urban amenities.

One of the familiar location models is the **gravity model**, which states that the interaction between two subareas is proportional to their activity levels, but inversely related to their spatial separation. **Reilly's law of gravitational attraction**, for example, is based on the concept of spatial interaction. One of the first retail models was constructed out of this theory. This model uses the number of business activities, people, store sales, area, and so forth as an index of size and the fundamental measure of attractiveness of a central place. Consider a household located at I' choosing between the shopping centers at A and B as shown in Figure 2.2, or the reverse situation where a shopping center I' is to be located to serve the population at A and B . In general, the markets captured from A and B are in the ratio

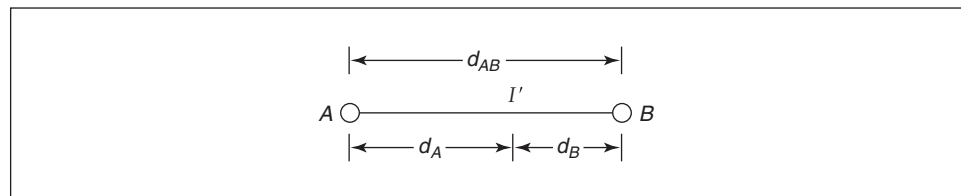
$$\frac{T_A'}{T_B'} = \frac{W_A}{W_B} \left(\frac{d_B}{d_A} \right)^2 \quad (2.1)$$

where W_A and W_B are the sizes of A and B , where T_A' , T_B' represent proportions of trade (percentage of sales for example) from I to A and B respectively, and d_B , d_A is the distance from B and A respectively, with $d_A + d_B = d_{AB}$.

From Equation 2.1 attractiveness of A and B with respect to point I' , when A and B are of equal size ($W_A = W_B$), can be represented as $T_A' d_A^2 = T_B' d_B^2$. Notice the appeal of A and B is a function of both distance away and sales volume. To locate a shopping center at I' equally appealing to both the population centers A and B , or to say it the other way, to find the point I' where a shopper is indifferent between shopping centers A and B , we set $T_A' = T_B'$ in Equation 2.1 and solve for d_B . In general, an equation can be derived that states the watershed trade area bounded between A and B , measured in miles (km) from B , is

$$d_B = \frac{d_{AB}}{1 + (W_A/W_B)^{1/2}} \quad (2.2)$$

Figure 2.2 BREAK POINT MODEL



Example

Let $d_{AB} = 36$ miles (57.6 km); $W_A = 92$ retail activities, $W_B = 90$ retail activities; then $d_B = 17.8$ miles (28.5 km) from location B according to Equation 2.2. ■

The Reilly model may be an acceptable approximation for such location decisions in rural areas where central places are rather distinguishable. In a more developed area, however, a large number of shopping centers and population centers are involved. The overlapping market areas will be too complex to be resolved by this idealized model. Another formulation of the gravity model was proposed by Lakshmanan and Hansen (1965). This model allocates retail dollars, determining the percentage of the population in subarea i that will go to the shopping center j to spend their money:

$$(\text{expenditure})_{ij} = (\text{expenditure})_i \frac{W_j / \tau_{ij}^\beta}{\sum_k W_k / \tau_{ik}^\beta}$$

where τ is the travel time and β is the positive exponent to be calibrated. This states that the total consumer retail expenditure of population in subarea i is allocated toward each shopping center j in accordance with the gravity formula. Notice travel distance d is replaced by time τ in this formulation. We will see more of this interchangeability between time and distance in subsequent discussions throughout this book. **Huff's probabilistic model** (1962) is yet another example of the gravity model, stating that the probability a consumer located at i will visit shopping center j is

$$\frac{W_j / \tau_{ij}^\beta}{\sum_k W_k / \tau_{ik}^\beta} \quad (2.3)$$

Example

Suppose there are two shopping malls 5 and 10 miles (8 and 16 km) away respectively, each with 800 and 300 thousand square feet (72 and 27 thousand m²) retail floor space. According to Huff's model, the probabilities a consumer will patronize these two malls are respectively

$$\begin{aligned} \frac{(800)(1/5^2)}{(800)(1/5^2) + (300)(1/10^2)} &= 0.08 \\ \frac{(300)(1/10^2)}{(800)(1/5^2) + (300)(1/10^2)} &= 0.92 \end{aligned} \quad (2.4)$$

assuming an exponent $\beta = 2$ (Dickey 1983). ■

Variants of location theory are found in literature on multicommodity flow as well as short-run and long-run equilibria of economic activities. **Multicommodity-flow models** describe the simultaneous allocation of population, employment, resources and finished products between places of supply and demand. In the short run, most economic activities, including the places of supply and demand, are fixed in location. In the long run, however, they could relocate themselves somewhere else corresponding to the rationing scheme of

the spatial price system. Short- and long-run multicommodity flows are often modeled by a generalized version of the gravity model and optimization models—subjects covered in Chapter 4.

C. Input-Output Models

Input-output models, developed by Leontief (1953), will be introduced with respect to two particular applications: local-impact studies and interregional-flow studies. As an example, local-impact studies reveal the possible changes in a single region. Interregional-flow studies, on the other hand, are to show the structural relationship between regions. The effect of an autonomous shock—such as the precipitous injection of basic employment into the study area as mentioned in economic-base theory—may be traced to, and through, the region under consideration. An essential part of an input-output model is an input-output table, which documents a set of economic multipliers similar to those found in economic-base theory. The input-output table (matrix) eventually gives rise to a set of simultaneous equations with production (or technical) coefficients (the multipliers) and activity variables. The set of equations can trace out, on a multi-sectoral basis, the implication of introducing a new industry into the study area (the autonomous shock). For example, if a new tourist trade is introduced into the area as a way to boost the local economy, what would be the implications on the economic activities associated with tourism such as the associated retail and entertainment industries? The set of simultaneous equations merely chain-up the sequence of effects together in a mathematical formulation through the use of a table or matrix where the rows are inputs (e.g., tourists) and the columns are outputs (e.g., retail sales). It can be thought of as a huge revenue/expenditure accounting system. The revenue side of the balance sheet shows how the output for each industry is distributed, and the expenditure side records for each industry the distribution-of-inputs per unit-of-output from all industries.

An example of such an input-output matrix is shown in Table 2.1 (Chapin and Kaiser 1979). Shown for a single region, the table records horizontally the output for each particular sector of the economy measured in terms of receipts from sales (of goods or services) to every other sector. Thus sector 1 may be the tourist industry, sector 2 may be retail, sector 3 entertainment, and sector 4

Table 2.1 EXAMPLE INPUT-OUTPUT TABLE

	Tourism sector	Retail sector	Entertainment sector	Household sector	Final demand
Tourism sector	\$30	\$20	\$30	\$25	\$105
Retail sector	60	20	80	30	190
Entertainment sector	10	40	60	50	160
Household sector	40	20	30	15	105
Charges against final demand	140	100	200	120	560

SOURCE: Chapin and Kaiser (1979). Reprinted with permission.

households. Households receive 25 million dollars during the current time period in wages as employees serving the tourist industry, the entertainment sector receives 30 million dollars from tourism, retail receives 20 million dollars, and the tourism sector spends 30 million on itself. Read vertically, the table shows input in terms of dollars spent on purchases in a particular sector from all other sectors. Thus local households as a whole spend 40 million dollars this time period on tourism, the entertainment industry spends 10 million dollars on the tourism industry as part of the intersectoral trade, and the retail industry spends 60 million dollars.

The final demand column records purchases by the tourism, retail, entertainment, and household sector—the dollar transactions after all intermediate processing and handling are completed. For example, tourists inject a total of 105 million dollars (first row sum) into the economy during this time period, divided among retail purchases, entertainment, and direct use of local labor. The charges against final demand in the bottom row are payments for tourism, retail trade, entertainment trade, and labor. Thus the fourth column (120 million) is the total wages paid to the household for supplying the labor for the remaining three sectors of the local economy, including the tourist industry, the third column is the total payment to the entertainment industry from other sectors and so on. These column totals are defined as the activity variables. To the extent that the row sums are not the same as column sums (or total purchases are not equal to payments) in Table 2.1, the final equilibrium values of these activities, taking the multiplier effects into account, are to be determined by the solution of a set of simultaneous equations.

From the dollar transactions in Table 2.1, production (or technical) coefficients are derived by dividing each input in a given column by the total of all inputs in the column. The resulting coefficients, shown in Table 2.2, are read by columns and indicate the cents-of-direct-inputs per dollar-of-output. Column 1 shows the input per dollar-value-of-output from each of all the other sectors supplying goods or services to sector 1. Thus the households contribute 29 cents toward the dollar on tourism, the entertainment sector contributes 7 cents, retail contributes 43 cents, and tourism pays itself 21 cents. The other columns show similar relationships for the retail, entertainment, and household sectors. The input-output technique, therefore, establishes a basic relationship between the volume output of any given industry in a region and the volume of input required in the production process from all other industries in this region. In this regard, the coefficients are equivalent to the labor force participation rate and population-serving ratio used in economic-base theory, except that the multipliers here are constructed out of dollar volumes rather than in terms of people. To the extent that intersectoral trade is governed by these multipliers aside from the seed activity (or autonomous

Table 2.2 PRODUCTION (TECHNICAL) COEFFICIENTS FOR A SINGLE REGION

	Tourism sector	Retail sector	Entertainment sector	Household sector
Tourism sector	0.21	0.20	0.15	0.21
Retail sector	0.43	0.20	0.40	0.25
Entertainment sector	0.07	0.40	0.30	0.42
Household sector	0.29	0.20	0.15	0.12

shock), the projection of the local economy, to be manifested in the final values of the activity variables, can only be determined following the four steps of economic-base theory, or alternatively solving the equivalent simultaneous equation set.

In the book, Chan (2005), more discussions of this Table can be found in the chapter on “Spatial Equilibrium and Disequilibrium.” The similarity between input-output theory and economic-base theory will be emphasized. Most important, the input-output model will be extended from the current intraregional version to an interregional version.

II. ECONOMETRIC MODELING: INTERREGIONAL DEMOGRAPHIC PROJECTIONS

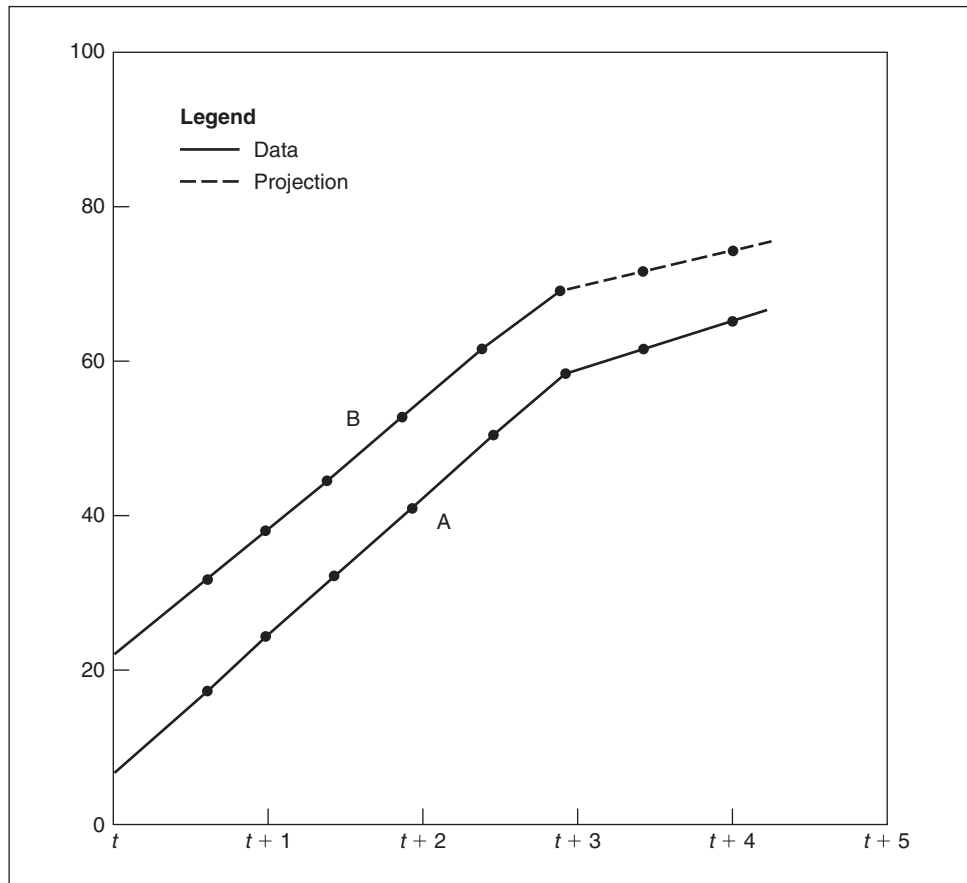
At the root of economic growth is population growth, for industrial wealth is nothing but a manifestation of human resources. An integral part of spatial economics is therefore the projection of population in a regional and interregional context. The demographic model is discussed here as a companion analysis to economic-base theory and input-output analysis. It also serves to illustrate economic theories, which are supplemental to classic economic theory in regional science. Three of the basic issues involved in demographic analyses are fertility, mortality, and migration. Fertility is the rate of childbirth in society. Mortality refers to the death rate in society. Migration is the population movement from one geographic location to another. Demographic analysis takes the net effect of fertility, mortality, and migration and predicts the growth or decline of population in the study area. The methods of analyzing demographic activities consist of population projection models, and matrix analyses of regional and interregional growth and distribution (Jha 1972). Population projection models are aggregate methods of extrapolating regional population growth from present trends using statistical techniques. The matrix analysis of population growth, on the other hand, is a more systemized method of projecting population growth, being more explanatory about the determinants of demographic activities.

A. Population Projection Models

Two of the key concepts used in the population projection models are comparative forecasting and extrapolation. **Comparative forecasting** is a very crude method and could be rather unreliable if performed carelessly. This forecasting method is performed by selecting two areas, *A* and *B*, which have behaved similarly in their demographic growth patterns. It is assumed that the two areas should develop similarly in the future, meaning that if *A*'s population increases at a certain rate, *B*'s population would increase at about the same rate. Notice that *A* can be a part of *B* geographically. Parallel attempts are made to establish population and employment growth rate for similar cities. (See the “Econometric Models” chapter in Chan [2005]).

Example

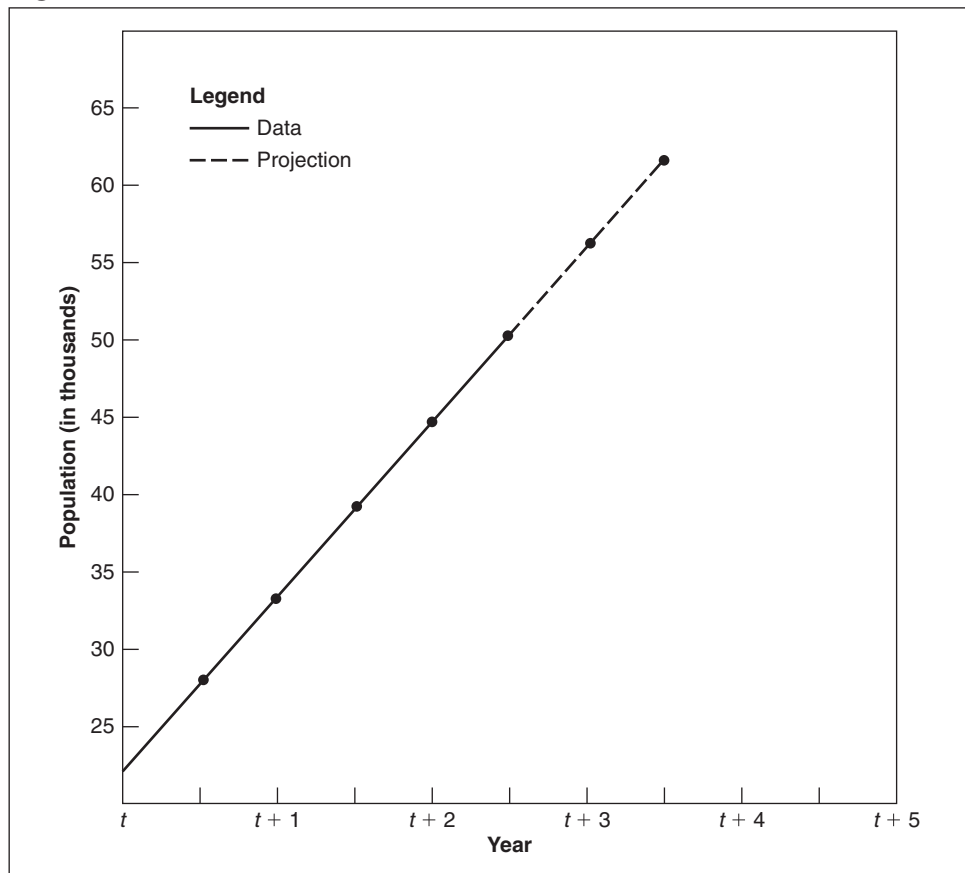
As shown in Figure 2.3, if the population growth of two areas *A* and *B* are similar in the past from t to $t + 3$, and if the population of *A* is known for the rest of the years from time period $t + 4$ to $t + 5$, we can have an idea of the population projection for area *B* for the corresponding years. In this method, we assume that the

Figure 2.3 POPULATION PROJECTION BY COMPARATIVE FORECASTING

demographics of one area follow the same profile as the other. This will be true even if there is a sharp decline in growth rate occurring around time period $t + 3$. ■

Extrapolation, on the other hand, uses statistical techniques to predict future population growth based on the trend in the same area in the past. This is the basic premise of almost all econometric models, in which the implicit assumption is that past trends prevail. It represents both the strength as well as the weakness of this type of model. It is a strength since the forecasting methodology is flexible and relatively easy to use. It is a weakness inasmuch as the underlying behavior of the study area is ignored, in preference for purely statistical correlations. The common techniques employed in comparative and extrapolation models are graphical, polynomial curves, ratio and correlation method, regression and covariance method, and inflow-outflow analysis.

1. Graphical Method. The graphical or manual technique consists of plotting points on a graph to show population growth predictions. In this method, past census data is used for plotting the graph of population versus time. Future population is obtained by extending the graph in the same way as the trend in the past. Thus in Figure 2.4, the population at $t + 5$ and $t + 6$ have been obtained by

Figure 2.4 GRAPHICAL PROJECTION OF POPULATION AT REGION C

extending the graph. Simple as it may look, graphic plots of data are an essential, indispensable first step in any econometric application. They allow the modeler to get a feel of the data and more importantly to formulate a hypothesis about the structural form of the model. Pairwise plots such as those shown in Figure 2.4 are options in almost all statistical analysis software. Actual projection may not be actually performed manually, but the trend indicated by the plot is a most important piece of information for the modeler.

2. Polynomial Method. The polynomial-curve technique is a generalization of the above concept. It is built upon the following linearized formula for each forecast increment Δt : $N(t + \Delta t) + \delta N(t) \Delta t$, where $N(t)$ is the base-year population, Δt is the forecast period (whether it be one year, five years or ten years.), and $\delta N(t)$ is the population increase per time period Δt .

Example

If for an area, the total population in base-year t is 4500 thousand and the annual increment has been 27 thousand, then the population in $t + 10$ will be equal to $4500 + (27)(10) = 4770$ thousand. ■

Polynomial curves are usually quite a bit more complex than the example shown above. For each time period, Δt , there exists a formal mathematical equation with a different increment as determined by the function $f(\Delta t)$: $N(t + \Delta t) = N(t) + f(\Delta t)$. Oftentimes, polynomial projections put more weight on present trends than past trends. One such weighting scheme is the **exponential smoothing technique** where the weight decays exponentially over the length of the elapsed time period, thus placing more value upon recent information. We will defer the details until the “Spatial Time-Series” chapter in Chan (2005), where formal projection methodologies will be discussed.

3. Ratio-and-Correlation Method. It might be possible that the population growth of the study area is related to the population growth of another area, or the region within which the area is located; or the population may be related to some socioeconomic factor such as employment of another area or the region. In this case, we use the ratio or coefficient of the relationship between the two areas for predicting future population, as shown in the following example.

Example

If the ratio of population at area A and any other socioeconomic factor at area B (including population) has been constant in the past years, then we can get the future area A population using this constant. Let $Z_B(t)$ represent the population or any other activity variable of area B at base year t , and suppose the ratio $N_A(t)/Z_B(t) = 0.8$.

If $Z_B(t + \Delta t) = 4000$ in the forecast year $t + \Delta t$,

$$\text{then } \frac{N_A(t)}{Z_B(t)} = \frac{N_A(t + \Delta t)}{4000} = 0.8$$

$$\text{or } N_A(t + \Delta t) = (4000)(0.8) = 3200 \blacksquare$$

In other words, the ratio-and-correlation method uses another activity variable to predict population growth, if population growth can be correlated with an identifiable activity variable at a different area via a constant ratio. The reader can imagine that an example can easily be constructed for the interregional input-output model where the population in a region, being the support labor force for an industry, is simply related to the employment level at the work region by the labor-force-participation rate. The gist of this method is straightforward. If $N_i(t)/Z_j(t) = \text{constant}$, then $N_i(t + \Delta t) = Z_j(t + \Delta t)(\text{constant})$. This model can be generalized to read

$$\frac{N_i(t + \Delta t)}{Z_j(t + \Delta t)} = f\left(\frac{N_i(t)}{Z_j(t)}, \frac{N_i(t - \Delta t)}{Z_j(t - \Delta t)}, \dots, \frac{N_i(t - n\Delta t)}{Z_j(t - n\Delta t)}\right) \quad (2.5)$$

where area i can also be area j ($i = j$), meaning that population and employment can be co-located in the same region. Here $f(\cdot)$ is a function showing how the constant can be determined by using historical information over n time periods. In the chapter on “Econometric Models” of Chan (2005), we will see how one can expand a great deal upon this very simple idea of ratio and correlation.

4. Regression and Covariance Analysis. This is one of the statistical calibration techniques widely used in population projection and for other activity variables as well. Here, population is taken as a dependent variable and another

activity or factor is taken as an independent variable. Usually a simple bivariate regression may be represented like this: $N = a + bX$, where X is any explanatory or independent factor, a and b are calibration constants that may be obtained by fitting the model to the regional data. The companion covariance analysis, or analysis of variance, measures the quality of the statistical fit of the model to the data.

Example

If the population of a state is associated with the increase in per capita income X , and a and b have been calibrated to be 2,095,000 and 1,062 respectively. Further suppose that the forecast-year per capita income in the state is 15,000, then according to the regression equation above, future state population is projected to be $(2,095,000) + (1,062)(15,000) = 36,880,000$. ■

In general, while the regression equation does not necessarily have to be linear to start out with, it is often reduced to the following linear form before calibration can be performed: $N = a + b_1X_1 + b_2X_2 + \dots$, where X_1 , X_2 and so forth are independent variables. The regression coefficients b_1 , b_2 and so forth are then calibrated for use in forecasting. Notice that the model assumes that the linear relationship between population and the independent variables will hold over time—very similar to the previous models, from comparative method to ratio-and-correlation method. The linearity assumption, and certain assumptions about the statistical distribution of the data, may impose restrictions on what is normally a very flexible modeling procedure. The technical aspects of regression and covariance analysis are discussed in Appendix 2 of this book.

5. Inflow-Outflow Analysis. The inflow-outflow analysis predicts the population of period $t + \Delta t$ into the future considering both the gain and loss of population in the area (termed **inflow** and **outflow** respectively.) The inflow is predicted by the equation

$$(\text{inflow}) = (\text{birthrate}) N(t) + (\text{in migration})$$

The outflow, on the other hand, is predicted by

$$(\text{outflow}) = (\text{death rate}) N(t) + (\text{out migration})$$

The population for the forecast year is predicted by combining the inflow and outflow results using the equation: $N(t + \Delta t) = N(t) + (\text{inflow} - \text{outflow})$. In summary, this method relates population projection to population growth, natural increase and decrease (due to birth and death respectively), and in-and-out migration via the following equation

$$N(t + \Delta t) = N(t) + \delta^N(\Delta t) + \delta^M(\Delta t)$$

where $\delta^N(\Delta t)$ is the natural increase or decrease in time period Δt , and $\delta^M(\Delta t)$ is the net migration during period Δt . Substituting and rearranging the terms, one can write $N(t + \Delta t) = N(t) + [b(\Delta t)N(t) + \delta^N(\Delta t)] + [d(\Delta t)N(t) + \delta^M(\Delta t)]$ where $b(\Delta t)$,

$d(\Delta t)$ are the birthrates and death rates during period Δt respectively, and $\delta N^r(\Delta t)$, $\delta N^l(\Delta t)$ are the in-and-out migrations during period Δt .

Example

If an area had a population of 4,500 for time period t , and the birthrate and death rate per capita are 2 and 1 percent respectively and the in-and-out migrations are 234 and 198 respectively for the forecast time increment, then the forecast population is $4,500 + (2(4,500/100) + 234) - (4,500/100 + 198) = 4,581$. ■

B. Interregional Growth and Distribution

Matrix representation of population growth and distribution is convenient for estimating the growth patterns of multi-regional populations. Two methods will be introduced here: **cohort survival** and **components of change**. The cohort survival method is a way to determine population growth. **Cohort**, for this purpose, is defined as a group of people born within a given time period. The fundamental concept of this analysis is: $\mathbf{N}(t + \Delta t) = \mathbf{G} \mathbf{N}(t)$, where the population at a future period $\mathbf{N}(t + \Delta t)$ is related to the current period t via a matrix \mathbf{G} , the growth matrix. For analytical purposes, the population is broken down into cohort age groups. The matrix takes into account the death rates for each age group and incorporates them as survival ratio at the main diagonal of the matrix. On the other hand, the birthrates for each of the age groups are represented in the first row of the matrix. For example, the birthrate for age groups under childbearing age is zero, and similarly for those over the childbearing age. However, each group within the childbearing age would have a certain birthrate, suggesting their capacity to reproduce. The matrix determines the populations, by age group, for the forecast year based on survival and birthrates. The matrix also ages the base-year population into older groups for the forecast year. A group of residents in the five-to-ten-year age bracket, for example, would transition into the ten-to-fifteen-year bracket if the forecast is performed for a five-year increment. In summary, the following equation set incorporates all the above elements in a matrix notation.

$$\begin{pmatrix} N_1(t + \Delta t) \\ N_2(t + \Delta t) \\ N_3(t + \Delta t) \\ \vdots \\ N_n(t + \Delta t) \end{pmatrix} = \begin{bmatrix} 0 & 0 & b_3 & b_4 & \cdot & \cdot & \cdot & b_{n-1} & 0 \\ s_{12} & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & s_{23} & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & s_{34} & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & s_{n-1n} & 0 \end{bmatrix} \begin{pmatrix} N_1(t) \\ N_2(t) \\ N_3(t) \\ \vdots \\ N_n(t) \end{pmatrix} \quad (2.6)$$

where b_i stands for the birthrate per person for group i , and s_{ij} stands for the surviving ratio of group i in group j .

Aside from birth-death considerations, the problem of interregional migration can be taken into account by using a migration matrix. This matrix is similar to that used to model the survival rates of cohort groups, except that net immigration and emigration rates are written in the main diagonal. Since the matrix is used to model interregional population movement alone, no birthrates are included. In the following matrix, where the row and column dimensions correspond to the different age groups, net interregional population migration is modeled:

$$\begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ m_{11} & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & m_{23} & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & m_{34} & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & m_{a-1\ n} & 0 \end{bmatrix}$$

The growth of a region is predicted by adding the birthrate, survival-rate, and migration-rate matrices, which produces a growth-rate matrix by age group

$$G = \begin{bmatrix} \leftarrow \bar{b} \rightarrow \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 \\ s_{12} & 0 & \cdot & \cdot & 0 \\ 0 & s_{23} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 \\ m_{12} & 0 & \cdot & \cdot & 0 \\ 0 & m_{23} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 \end{bmatrix} \quad (2.7)$$

Example

A simple numerical example would illustrate these matrices. Consider three age groups: 0- to 20-year-olds, 20- to 40-year-olds and 40- to 60-year-olds. These hypothetical matrices can be written:

$$G = \begin{bmatrix} 0 & 1.5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0.9 & 0 & 0 \\ 0 & 0.8 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \end{bmatrix} \quad (2.8)$$

where the childbearing cohort group is defined as those 20 to 40 years old. We specify that 9 out of 10 people survive from the 0- to 20-year group to become 20- to 40-year-old adults. Ten percent more people in the 20- to 40-year-old group migrate into the area over 20 years—the length of the forecast period—and so on. Summing these matrices, we have the net growth matrix

$$G = \begin{bmatrix} 0 & 1.5 & 0 \\ 1.0 & 0 & 0 \\ 0 & 0.9 & 0 \end{bmatrix}$$

If the base-year population in all age groups is 10,000, the forecast population distribution (in thousands) would be

$$\begin{pmatrix} N_1(t + \Delta t) \\ N_2(t + \Delta t) \\ N_3(t + \Delta t) \end{pmatrix} = \begin{bmatrix} 0 & 1.5 & 0 \\ 1.0 & 0 & 0 \\ 0 & 0.9 & 0 \end{bmatrix} \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix} = \begin{pmatrix} 15 \\ 10 \\ 9 \end{pmatrix} \quad (2.9)$$

It is predicted, therefore, that in 20 years more young people than older people will be living in the study area. More precisely, there will be 15 thousand 0- to 20-year-olds, 10 thousand 20- to 40-year-olds, and only 9 thousand 40- to 60-year-olds. ■

C. Interregional Components of Change Model

Predicting interregional population is basically the same as predicting regional population. The major differences are that instead of breaking down by age groups, we stratify by specific regions, such as the East versus West Coast. This basic concept is still used:

$$N(t + \Delta t) = N(t) + (\text{births}) - (\text{deaths}) + (\text{migrants})$$

Symbolically, the components of change model may be stated in scalar terms for each region i as

$$\begin{aligned} N_i(t + \Delta t) &= N_i(t) + b_i(t)N_i(t) - d_i(t)N_i(t) + m_i(t)N_i(t) \\ &= [1 + b_i(t) - d_i(t) + m_i(t)] N_i(t) \\ &= g_i N_i(t) \end{aligned} \quad (2.10)$$

where b , d , and m are birth-, death and net migration rates. For example, the crude birth-, death and net migration rates from Table 2.3 give rise to the growth rate $g = 1 + 0.1315 - 0.0473 + 0.0865 = 1.1707$. These are called crude because they are simply the births, deaths, and net migration over the period 1955–60 divided by the 1955 base-year population in California, without taking into consideration migration from/to the rest of the United States or any place else. In fact, proper estimation of these parameters is a subject of interest in real world applications. Chan (2005) elaborates on this topic in the “Bifurcation and Disaggregation” chapter. (Software and data, under the YI-CHAN folder, are also included on the CD/DVD attached to this book to illustrate the estimation procedure.) Usually, population, births, deaths, and migration are expressed in matrix forms, where the row and column dimensions correspond to the number of regions being modeled. The following model shows a two-region example in which the internal births, deaths, and interregional net migration are analyzed.

$$\begin{pmatrix} N_1(t + \Delta t) \\ N_2(t + \Delta t) \end{pmatrix} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} b_1(t) & 0 \\ 0 & b_2(t) \end{bmatrix} - \begin{bmatrix} d_1(t) & 0 \\ 0 & d_2(t) \end{bmatrix} + \begin{bmatrix} 0 & m_{21}(t) \\ m_{12}(t) & 0 \end{bmatrix} \right\} \begin{pmatrix} N_1(t) \\ N_2(t) \end{pmatrix} \quad (2.11)$$

or in matrix notation $N(t + \Delta t) = (I + B - D + M) N(t) = G N(t)$.

Table 2.3 CALIFORNIA AND THE REST OF THE UNITED STATES (1955–60)

Region	1955 Pop	Birthrate	Death rate	Migration rate
Calif	12,988,000	0.1315	0.0473	0.0865 (~US to Calif)
Rest of the US (~US)	152,082,000	0.1282	0.0488	−0.0074 (Calif to ~US)

Example

From the data in Table 2.3, the growth matrix is the sum of the identity, birth, death, and migration matrices, where California is row/column 1 and the rest of the United States is row/column 2 of such matrices:

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1315 & 0 \\ 0 & 0.1282 \end{bmatrix} - \begin{bmatrix} 0.0473 & 0 \\ 0 & 0.0488 \end{bmatrix} + \begin{bmatrix} 0 & 0.0865 \\ -0.0074 & 0 \end{bmatrix} = \begin{bmatrix} 1.0842 & 0.0865 \\ -0.0074 & 1.0794 \end{bmatrix} \quad (2.12)$$

The 1960 population in California and the rest of the United States can then be computed as

$$\begin{pmatrix} N_1(1960) \\ N_2(1960) \end{pmatrix} = \begin{bmatrix} 1.0842 & 0.0865 \\ -0.0074 & 1.0794 \end{bmatrix} \begin{pmatrix} 12,988 \\ 152,082 \end{pmatrix} = \begin{pmatrix} 27,236 \\ 164,061 \end{pmatrix} \quad \blacksquare \quad (2.13)$$

The discussion on interregional demographic model gives the reader a flavor of the basic algebra found in similar model structures as the interregional input-output model. It serves not only to introduce econometric modeling, but also to generalize to a multi-regional level a key projection concept introduced earlier in this chapter.

III. ECONOMIC CONSTRUCTS FOR COST-BENEFIT ESTIMATION

The previous sections have been devoted to the economic and econometric techniques of prediction where future activities, such as the local economy, are projected. In this section, we will concentrate on the methods of evaluation, in which a location or land use policy is analyzed or evaluated with respect to its cost and benefits. There are three economic concepts that are important to cost-benefit estimation: equity, efficiency, and externality. **Equity** is a very precise concept in economics since it connotes the distribution of income and social benefits. An example may be the equal accessibility of all segments of the population to such public services as school and recreation (Marsh and Schilling 1994). Equity can be achieved through the natural market forces, governmental intervention, or through public services and transfer payments. The price system may sometimes be inadequate to effect an equitable distribution of goods and services; it may then be necessary to subsidize schools in a less affluent neighborhood in order to render education opportunities for all.

Efficiency, in our context, means the least costly distribution of resources over space for the production of goods and services. An efficient urban structure, for example, is to have complementary goods and services to be clustered together, whereby transportation costs are minimized.¹ Such a clustered development may mean the sacrifice of some open space that is sometimes highly valued. Efficiency, therefore, is not necessarily the only objective of urban planning; other factors need to be considered at the same time.

Externality, for the purpose of the current discussion, refers to the effects of a project other than those measured by the economic price system. In the provision of open space above, transportation cost does not accurately represent the price for the distribution of open space around the city, meaning that a precise, quantifiable price measure of the value of open space to an inhabitant is not easily obtainable. Economists have a well-defined concept about price theory, and they recognize that certain effects cannot be measured by price, including the positive benefits of open space and the negative benefits of air pollution. However, in a comprehensive accounting system, we may like to impute a cost to the community for the deprivation of open space, or the onset of pollution, both of which may be incurred in the industrial production process. This imputed cost is an example of an externality (Dahlman 1988).

Having been equipped with these basic concepts, we are prepared to examine two sets of methodologies for estimating costs and benefits. The first is **shift-share analysis**, which illustrates a technique to measure equity in a spatial context. The second is **theory of land values**, which is included here to verify the concept of efficiency.

A. Shift-Share Analysis

Shift-share analysis is a technique to divide the change in a socioeconomic measure into two or more components. For example, the population growth in an area is attributable to both the regional growth pattern and the peculiarity of the area itself. This technique can be used to measure the distribution of benefits: for instance, which subarea in the study area will receive less than its equitable share of regional growth and which will receive more. Rather than assuming a constant trend and a constant share of the regional economic activities, shift-share analysis tries to explain the change in the activity level in a particular subarea by two components. The first component is an average activity change corresponding to an aggregate regional change, while the second component is the difference between the average and actual changes in a subarea. This can be expressed by the following equation:

$$(\text{subareal change}) = (\text{regional average change}) + (\text{competitive change})$$

For example, an urban area grows 10 percent over a five-year period, and two of its zones *A* and *B* grow by five percent and 12 percent respectively. Zone *A* is at a competitive disadvantage of five percent below while zone *B* is at an advantage of two percent above the regional average, even though both are influenced by the overall regional growth.

A general expression of shift-share analysis can be written for activity *k* in subarea *i*:

$$\Delta Z_i^k = \delta Z^k + \delta Z_i^k = \frac{\Delta Z^k}{Z^k(t)} Z_i^k(t) + \delta Z_i^k \quad (2.14)$$

which states that the total change of activity *k* in subarea *i* is due to subareal change of activity *k* at the regional rate, adjusted for site-specific change at the local level. Shift-share analysis is therefore a simple concept of splitting up the change in activity from time period *t* to *t* + 1 into two functional components.

The first component indicates the norm for the region as a whole and the second the subareal deviation from the norm as mentioned. Notice that the competitive component is introduced to measure the change in a subarea relative to the regional average—showing the relative attractiveness of the subarea for the particular activity under consideration.

The competitive component of change in activity k for subarea i , δZ_i^k , can again be broken down into two components: the difference between subareal change and the regional overall growth in sector k .

$$\delta Z_i^k = Z_i^k(t) \left[\frac{\Delta Z_i^k}{Z_i^k(t)} - \frac{\Delta Z^k}{Z^k(t)} \right] \quad (2.15)$$

Putting it altogether, we can see that δZ^k in Equation 2.14 defines the change in importance of industrial sector k in subarea i over the time period, or the shift component. Equation 2.15, on the other hand, defines the increase or decrease in activity k due to the relative competitiveness of subarea i vis-a-vis other subareas, or the share component. This accounts for the name shift-share analysis.

Example

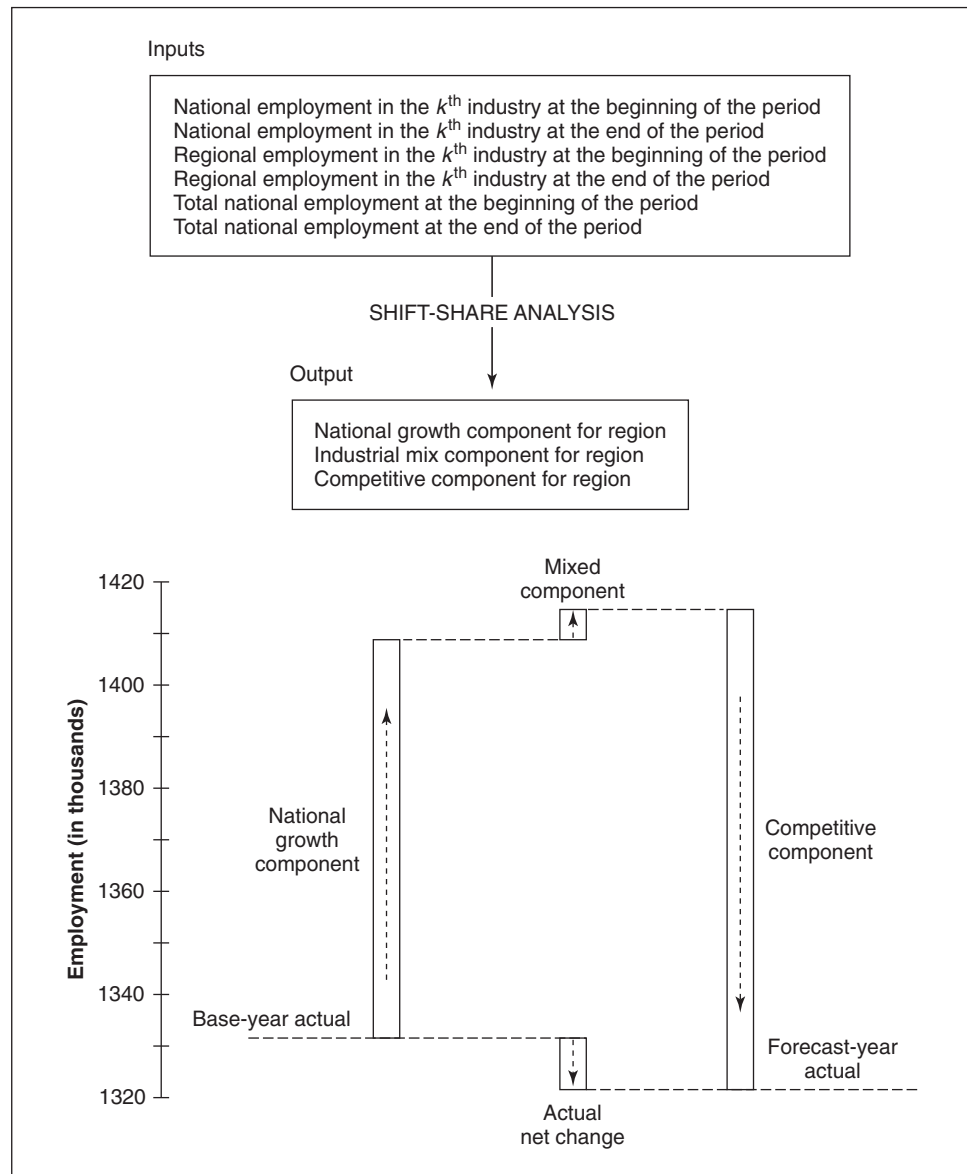
During the past five years, subarea i 's manufacturing (M) sector grew less rapidly than did the region by 1.6 percent. Its commercial (C) sector, in contrast, had a growth rate that exceeded that of the region's by 3.8 percent. Regional manufacturing and commercial growth rates are given as 0.276 and 0.402 respectively (i.e., 27.6 percent and 40.2 percent), and the current subareal manufacturing and commercial activity levels are \$280,000 and \$180,000 respectively. Assuming a constant shift, what is the value of manufacturing and commercial trade in a projected time period?

To answer this question, we add the national growth rate to the subarea's growth rate and multiply the result by the subarea's current sectoral activity level according to Equation 2.15, yielding the projected manufacturing and commercial levels as requested:

$$\begin{aligned} \delta Z_i^M &= (-0.016 + 0.276)280 = 72.8 \\ \delta Z_i^C &= (+0.038 + 0.402)180 = 79.2 \end{aligned} \quad (2.16)$$

In this shift-share example, the first term in Equation 2.14 disappears since we assumed constant shift (Krueckeberg and Silver 1974). ■

Figure 2.5 illustrates another example in the relationship between a regional economy and the national economy where all three components are present: national growth component, industrial mix component, and the competitive component. It shows the input data required to estimate each of these components, as well as a graphic plot of a numerical example for regional employment. Thus the drop in regional employment from 1332 to 1321 thousand is explained in terms of these components. The concepts presented in shift-share analysis, while simple, are not readily used in the field, since we never discussed how the growth rates are actually derived beyond the schematic as illustrated. Chan (2005) shows in his "Spatial Equilibrium and Disequilibrium" chapter that implementation potentials can be enhanced by including this concept within the interregional version of input-output analysis.

Figure 2.5 EXAMPLE APPLICATION OF SHIFT-SHARE ANALYSIS

B. Theory of Land Values

Having completed our discussions on equity measurement, let us now turn to the concept of efficiency and illustrate it through the theory of land values. Land value is subject to the market forces of supply and demand and highly related to location and transportation costs. An improvement of the transportation system, such as new highway or subway construction, could affect land value significantly. Dorau and Hinman, as far back as 1928, suggested tracing land value to three additional explanatory variables: land income, rate of capitalization, and

direct satisfaction from land ownership. Land income includes mortgages as well as the rent collected from tenants on the property, and in general the usefulness of the land corresponding to the various services it can render. While it may be obvious that land value depends on how potential income can be obtained from the land, it needs to be pointed out that such income includes not only those from the current time period, but also the forthcoming periods. This means that the rate of capitalization, such as interest rates, risk, and other investment preferences, are involved. The last explanatory factor—direct satisfaction from ownership—needs little explanation. It pertains to the personal rewards that are not measured by the monetary system.

Thus it can be seen that in a cost-benefit analysis, if land value is the primary measure of benefit, there are a variety of means to effect the change in land value, each of which would probably incur a cost. Improving accessibility by building highways, for instance, is a way among many others. The theory of land values helps to explain such a cost-benefit relationship, and in a practical sense, contributes toward model building. Aside from the above observations, there are several economic phenomena that are useful for model building as well. It is observed, for example, that land value or land rent declines with the distance from the central business district. The further one goes away from the central city, the lower the land value. Land rent and transportation costs are complementary. Thus in a hypothetical, circular city, the land values can be viewed as a cone in three dimensions (see Figure 2.6). If one wishes to live in the central city, the land rent is at a peak, but the transportation costs are at a minimum. On the other hand, if one locates at the fringe of the city, the land rent will be low, but the transportation cost will be high. You can either pay a high rent and be accessible, or you can pay a low rent and be comparatively inaccessible, hence having to pay more on transportation costs. Land rent is affected by transportation in another way. In the case of Philadelphia and other cities with a radial highway system, the development follows along the freeways in a finger-like manner. Suppose one adapts Burgess's classic concentric zone structure to an urban area consisting of contours of land value in rings around the city center. After a freeway is built, the development would tend to align itself along the freeway, stretching out the rings as indicated in Figure 2.7. In this case, Burgess's theory merges with Hoyt's sector theory,² which suggests that there are modifications to the Burgess's concentric rings to reflect transportation corridors that induce suburban development along the corridors.

Figure 2.6 LAND RENT AND TRANSPORTATION COST

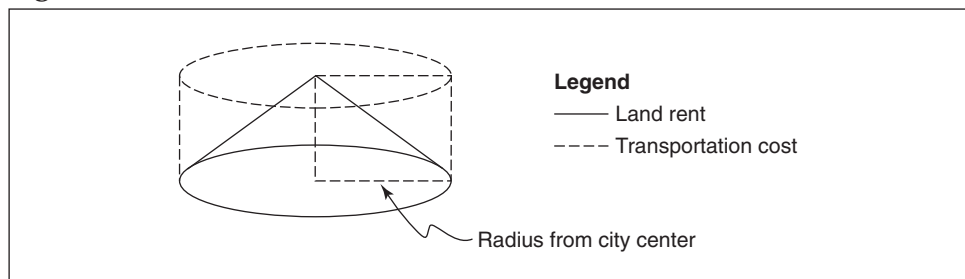
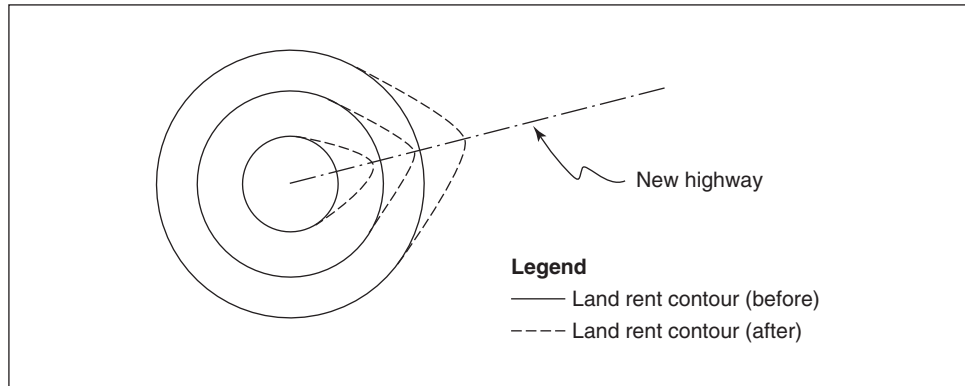


Figure 2.7 EFFECT OF TRANSPORTATION ON LAND USE

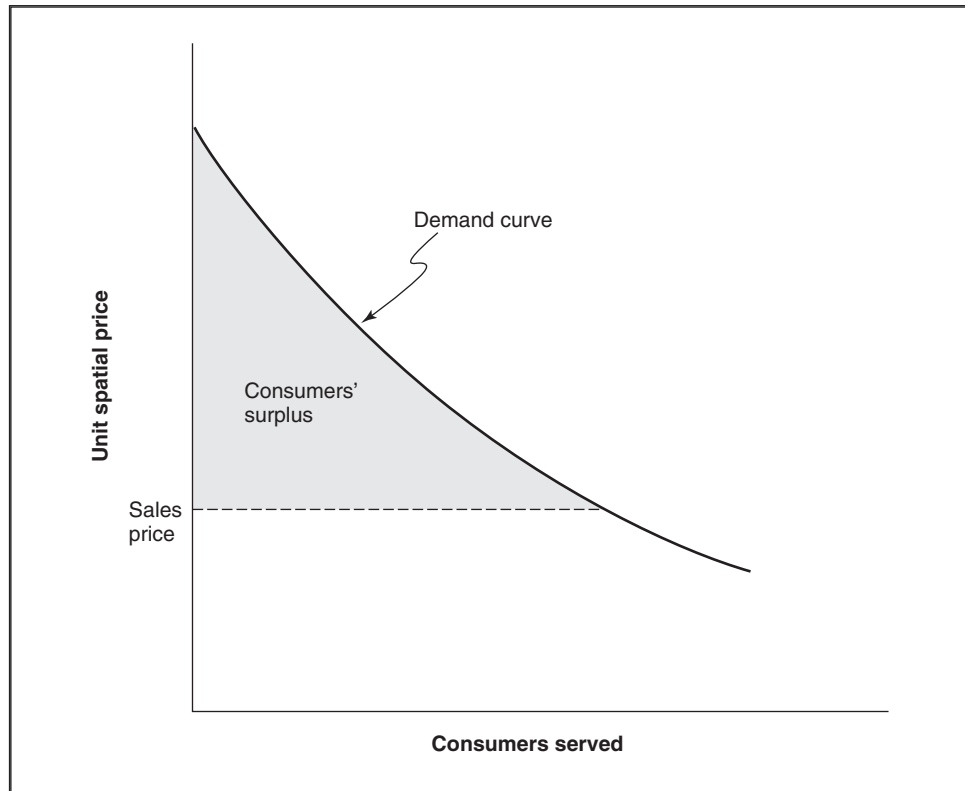
C. Consumers' Surplus

When economic efficiency is of concern, a valuation measure in spatial choice is **consumers' surplus**. The consumers' surplus is defined as the difference between what consumers might be willing to pay for a location and what they actually pay. As shown in Figure 2.8, the consumers' surplus is the area between the demand curve and the spatial price. Since the demand function expresses the users' indifference between the utility of a location and money, it can be considered as an expression of the utility of locations in terms of prices. The consumers' surplus, which is expressed in monetary units, is then a measure of the utility provided to the consumer minus the cost of production, which is reflected in the sale price to some degree. Maximization of consumers' surplus is then a close proxy of the maximization of the economic utility of the consumers. The evaluation of projects through a consumers' surplus analysis is widely, although generally only implicitly, used for large-scale public facilities. It is the only effective means of estimating economic benefits when the public facilities are so large as to effect more than marginal changes in prices.

To estimate the change in consumers' surplus brought about by any project, it is necessary to know both the price and the scale of the facility built before and after the project is completed. Figure 2.9 shows the change in consumers' surplus before and after a facility expansion from \bar{P}_{bef} to \bar{P}_{aft} , which increases the number of consumers served from V_{bef} to V_{aft} . Algebraically, this change can be approximated by the trapezoid rule:

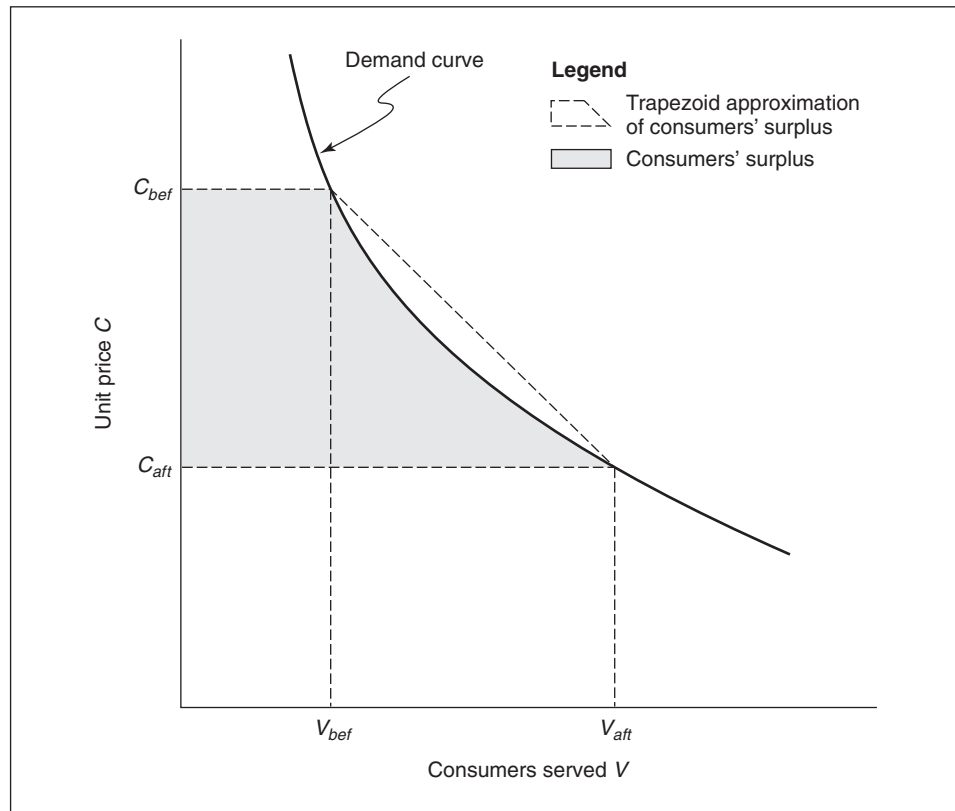
$$\frac{1}{2}(C_{bef} - C_{aft})(V_{bef} + V_{aft}) \quad (2.17)$$

Measurement of the equilibrium price C can be difficult when the project is large enough to shift the demand curve by causing an income effect. Such an income effect is illustrated in Figure 2.10, where the tradeoff between housing and transportation is considered.³ The effective increase in income caused by a price reduction on a major facility shifts the point of maximum utility from U_{bef}^* to U_{aft}^* . The increase in income thus results in an increased demand for both transportation and housing. The income effect of a price change is only significant when major

Figure 2.8 CONSUMERS' SURPLUS ILLUSTRATION

expenditure items are involved. For most families in the United States, these would be transportation, housing etc. Price changes on these items can change the level of consumption. Increased rent or housing costs could, for example, decrease the demand for travel. In developing countries, investments in basic infrastructure such as transportation, housing, and power can, by decreasing the cost of these items, significantly increase the effective income (I') of the inhabitants.

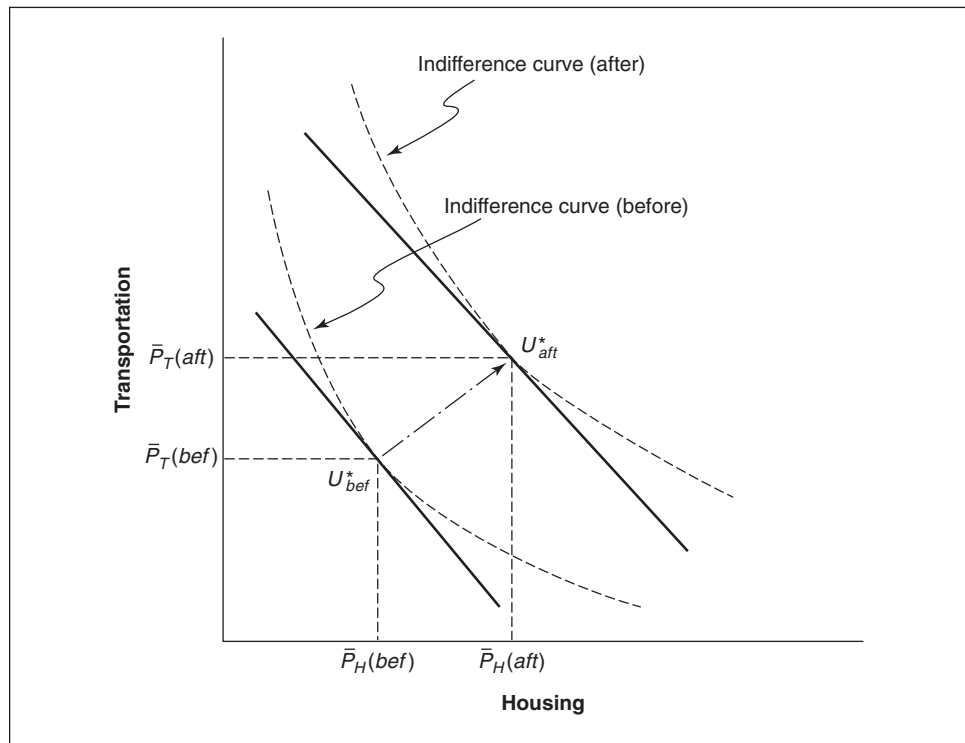
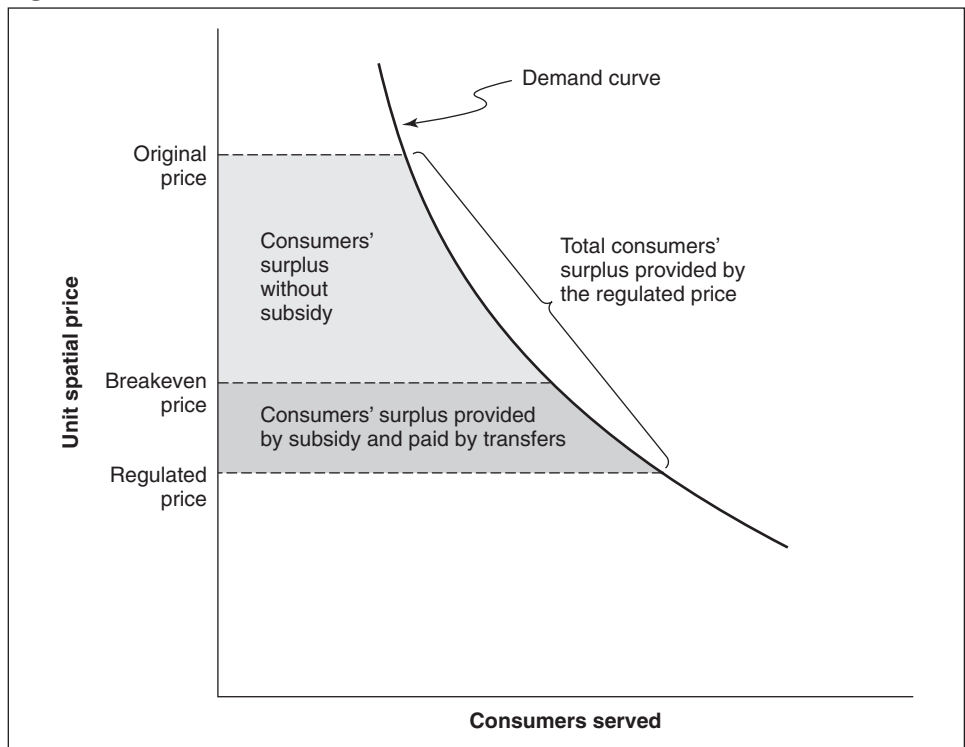
When income effect is involved, knowledge of the income elasticity of demand $\left(\frac{dV}{V} / \frac{dI'}{I'}\right)$ is required in order to estimate the final price C_{aft} along the same demand curve. Equation 2.17 still provides a satisfactory, although more approximate, means of calculating consumers' surplus. Chan (2005) illustrates this calculation in his "Including Generation and Distribution" chapter, where he estimates the economic value of state parks. (The software that performs such calculation is included on the attached CD/DVD under the STATEPRK folder.) In calculating consumers' surplus, the analyst must be careful to reckon with the effects of manipulations of the prices through a deliberated pricing policy. In systems that are publicly owned, it is possible and sometimes desirable to set prices that cover more or less than the total costs. Hydroelectric power in the western United States, for example, was subsidized below average cost to promote development. Unless the subsidies are deducted, this policy clearly increases consumers' surplus over what it might be if full cost of the service were charged. Figure 2.11 shows the total consumers' surplus made up of that part by the market mechanism and the other

Figure 2.9 CHANGE IN CONSUMERS' SURPLUS

part by regulation. Such changes in consumers' surplus, effected by setting the prices of services different from their costs, are not without expenditure. The changes are indeed transfer payments that must be made up by subsidy, either from taxes or from profits in some other part of the system and deducted from the final consumers' surplus calculations of the project.

IV. UTILITY THEORY

Utility theory is a common economic concept to explain location choice and decision among alternatives in general. A view of utility functions may be developed in the following way. Each household is confronted with a choice between n different expenditures, including savings or dis-savings, within an income budget. This can be expressed by the following equation where p_i and x_i refer to the price and quantity of the i^{th} expenditure: $I' = \sum_{i=1}^n p_i x_i$. On the other hand, the household derives a certain amount of satisfaction from the quantities of each commodity it purchases, and this degree of satisfaction, when added up, provides a total utility. This utility may be expressed as a function of the vector of purchases of

Figure 2.10 THE INCOME EFFECT**Figure 2.11** SUBSIDY AND TRANSFER PAYMENT

commodities and services \mathbf{x} : $v = f(\mathbf{x})$; but this expression is vacuous until we specify the form of the function $f(\mathbf{x})$. One may assume, for example, that it could be linear:

$$v = \sum_{i=1}^n w_i x_i \quad (2.18)$$

This says that utility is the weighted sum of the purchases. This turns out to be not a very satisfactory idea because if a household tried to maximize its utility under this simple form, the whole budget would be spent on the commodity or service for which w_i/p_i was a maximum. Thus if the weight on travel was high and transportation cost was low, a family might spend its entire income on travel, which is somewhat absurd.

It would not help very much if we retain the linear model of Equation 2.18, but placed a requirement on the minimum consumption of each x_i . This would result in every commodity being consumed at its minimum level with the exception of the most cost effective one. A more complicated model can easily be devised in which various needs are each satisfied by a linear combination of commodities, and minimum values are set for the satisfaction of each need. This model is still unrealistic in that the minimum level of needs has to be set exogenously. Normally within the household, choices are made between the levels of satisfaction of various broad classes of needs—the need for housing, accessibility, non-housing, and non-location goods and services. Any linear model would force us to make decisions about these tradeoffs outside the model.

What makes tradeoff and consumption both possible and necessary is the fact that, for most goods, increasing quantities provide increasing satisfaction, but at a decreasing rate. Thus if twice the space is available to a household by moving further away from the city, the increased space may not double the housing satisfaction. In some cases, it might even decrease it. If we assume that increasing amounts of a commodity always add something to a household's utility, or at least never subtract from it. Suppose we also assume that the increase in satisfaction for each additional unit of a given commodity is diminishing, we have familiar economic statements about utility functions which are usually expressed mathematically:

$$\begin{aligned} \partial v / \partial x_i &\geq 0 & i = 1, 2, \dots, n \\ \partial^2 v / \partial x_i^2 &\leq 0 & i = 1, 2, \dots, n \end{aligned} \quad (2.19)$$

An example function is

$$v = \sum_i a_i \ln x_i \quad (2.20)$$

or alternatively

$$v = \prod_{i=1}^n a_i x_i \quad (2.21)$$

A form of the utility function corresponding to these two is extremely useful for our discussion here because we are dealing with commodities which are, in the western culture, absolutely essential. Every family must have housing, access to employment, and other commodities such as food and clothing. If one of these commodities is reduced to zero in Equation 2.21, the level of utility falls to zero. A utility function of this type leads to tradeoffs that give adequate weight to extreme deprivation of any of the essential commodities of life. While Equations 2.20 and 2.21 are useful utility-function forms, alternative approaches exist to quantify a decision maker's values. In Chapter 5 the multi-attribute utility theory will be introduced, which is based more on behavioral grounds.

A. Estimating Bid-Rent via Utility Function

Before utility can be measured, the terms of the utility function must be defined. Part of the satisfaction from a particular residential location may be associated with the accessibility to work and/or recreational facilities in an area. Another may be connected to the availability of schools or pleasantness and quiet of the community. Let us now see how these are actually being quantified. First, we stratify the population by income, family size, and other socioeconomic factors, not only to detect different behaviors, but also to be sure that we are dealing with relatively uniform levels of housing and related expenditures. In the discussion that follows, it should be understood that income is fixed at a class mean, or at least falls within a relatively narrow range as a result of the stratification of individual households.

Alonso (1970) has the idea of measuring utility with reference to income, whereby the utility function takes into consideration the total available income. In a family's budget, let us define M' as the non-location expenditures, which include items such as food, clothing, and education. M' also includes savings at a bank. Another expenditure is rent (r), which includes mortgage payments, rent, and utility bills. Then we have transportation cost represented by T . Collectively r and T are referred to as location expenditures. These budget components can be broken down further, but the way we are doing it now satisfies our purpose. All these expenditures must fit into the budget I' : $I' = M' + r + T$, which says certain parts of the income go to location and another to non-location expenditures. The simple equation above also underlines the complementary relationship between transportation outlay and rent, as covered earlier in this chapter when we discussed land rent theory.

We will now assume a particularly simple form of the utility function referenced as Equation 2.20:

$$v = \ln M' + \alpha_1 \ln H + \alpha_2 \ln A + \alpha_3 \ln C' \quad (2.22)$$

Here, M' stands for the consumption of all non-location goods as discussed above, while H , A , and C' stand respectively for the expenditure on providing housing, accessibility, and community amenities. In Equation 2.22, α_1 , α_2 , and α_3 are coefficients defining the relative importance of housing, accessibility, and amenities. We now introduce a basic assumption of overriding importance, whose application to this problem is due to Alonso (1964, 1970). We assume that

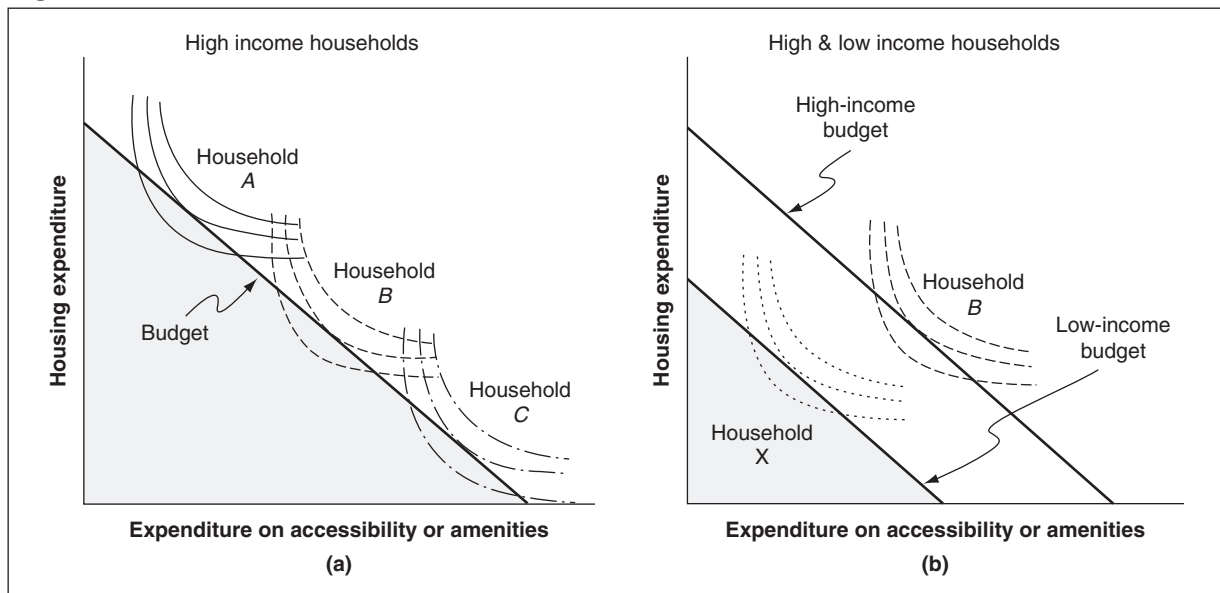
for a particular set of households of homogeneous tastes, utility is uniform wherever they are located in the metropolitan area. We cannot, of course, be sure that by defining homogeneous socioeconomic groups, we have actually defined groups whose preferences in the housing market are also homogeneous. Given some uniformity in tastes, however, the assumption of equal utility is based on elementary economic considerations. If the utilities being enjoyed are in fact not equal and if there are locations in which a particular group could enjoy a higher utility, members of that group will bid up the price of land and housing at that location. The higher cost of the housing package in this preferred area will, via the budget constraint, reduce the amount of money available for purchase on non-location commodities and thus reduce the level of utility enjoyed. Given freedom to move in search of better housing opportunities, this type of bidding will raise demand in some locations and lower it in others to the point where all utilities for this group have been equalized. This implies that there is a competitive equilibrium and the assumption for freedom to move is again important in achieving this equilibrium. See household groups *A*, *B*, and *C* of a high income class trading off their preference between housing, accessibility, and amenities expenditures in Figure 2.12(a). This contrasts with two households *B* and *X* in a high and low income class respectively shown in Figure 2.12(b).

Given that the utilities of any particular locating group are fixed at any particular point in time, the v which appears in Equation 2.22 is a constant, and we redefine it as

$$v = \ln I' + \ln F \quad (2.23)$$

Since we are dealing with a homogeneous income group, $\ln I'$ is a constant and F is an arbitrary constant whose role will appear below. If we now substitute

Figure 2.12 UTILITY FUNCTION AND BUDGET



SOURCE: Adapted from Yeates and Garner (1980). Reprinted with permission.

$M' = I' - r - T$ and Equation 2.23 in Equation 2.22 and rearrange terms, we arrive at the following expression:

$$\ln ([I' - r - T]/I') = \ln F - \alpha_1 \ln H - \alpha_2 \ln A - \alpha_3 \ln C' \quad (2.24)$$

This is an estimating equation which can be empirically tested and which expresses the proportion of non-location expenditures undertaken by each family as a function of housing, accessibility, and community amenities in each location. This equation has two essential properties. First, all the variables in it can be observed for a number of different household classes in a number of different locations, and consequently it can be determined. The level of non-location expenditures can be estimated from this equation and then, since I' and T are known, the rent which would be offered can be estimated using this equation.

We will show how this important procedure can be achieved. If we exponentiate Equation 2.24, we get $(I' - r - T)/I' = FH^{-\alpha_1}A^{-\alpha_2}C'^{-\alpha_3}$. Rearranging terms, we can isolate rent on the left-hand side of the equation. We show this value of rent as an estimated value:

$$r = I' - T - I'FH^{-\alpha_1}A^{-\alpha_2}C'^{-\alpha_3}$$

This is equivalent to the form of the budget equation $r = I' - T - M'$. These values of r are bid-rents discussed by Alonso in his development of the theory of location behavior. Expressing $\ln (1 - [r + T]/I')$ in Equation 2.24 in series, and recognizing that $(r - T)/I'$ is a fraction, an approximation can be made only by taking the first term of the series expansion⁴:

$$\ln \left(1 - \frac{r + T}{I'} \right) \approx - \frac{r + T}{I'} \quad (2.25)$$

This says that our dependent variable is approximately equal to the (negative) fraction of income spent on rent and transportation combined. This is analogous to the dependent variable of many of the housing market analyses: the rent-income ratio.

Notice the location expenditure is small compared to the rest of the budget for a majority of the population. The fraction of income spent on location expenditures can be estimated by this simple formula; it serves as an approximation for the dependent variable in the Equation 2.24. The above analysis indicates that there is substantial uniformity in the behavior among groups that have been defined on socioeconomic grounds. This behavior can be characterized through utility functions of a fundamentally simple nature. Data are available in the census and elsewhere for providing values for these estimates. All of the relevant variables that we suggested on a priori basis turn out to be statistically significant. The uses to which this analysis can be put must be discussed in conjunction with modeling the market clearing mechanism for housing. (See the Herbert-Stevens model in Chapter 4.)

B. Minimum-Cost Residential Location

Alonso's model of residential location would hold that households are located to minimize the cost of housing and travel. For a monocentric metropolis, this cost is expressed simply as $C(d) = H + r(d) + a'Vd/l'$, where $C(d)$ is the total location cost as a function of distance from the metropolitan area's center, the land area desired for the parcel of land is assumed constant, $r(d)$ is the cost of a unit-of-land as a function of location, a' is the unit cost of commuting (cost per unit-of-distance-traveled), d is the location's distance from the workplace at the metropolitan center, l' is the real discount rate on commuting trips due to such modern day conveniences as telecommuting, and V is the number of one-way commuting trips taken per year (Lund and Mokhtarian 1994).

Since households are assumed to minimize this cost in their location decisions,

$$\dot{C}(d^*) = \dot{r}(d^*) + a'V/l' = 0 \quad \text{or} \quad \dot{r}(d^*) = -a'V/l' \quad (2.26)$$

where the derivatives are evaluated at d^* , the least-cost residential location. Inasmuch as land prices tend to decrease with distance from the metropolitan center, $\dot{r} < 0$. So long as this relationship holds and to the extent that telecommuting lessens the number of work trips per year ($V_1 < V_0$), telecommuting is associated with a more gentle land-rent gradient:

$$\dot{r}(d^*)V_0 < \dot{r}(d^*)V_1 < 0 \quad (2.27)$$

Assuming that land prices follow a conventional exponential decay, then $r(d) = r_0 \exp(-Kd)$, where r_0 is the land price at the metropolitan area center and K is a decay constant. Therefore,

$$\dot{r}(d) = -r_0 K \exp(-Kd) \quad (2.28)$$

Combining Equations 2.26 and 2.28 yields $r_0 K \exp(-Kd^*) = a'V/l'$. This results in the least-cost residential location

$$d^* = (l'/K) \ln [l'r_0 K/a'] - (\ln V)/K \quad (2.29)$$

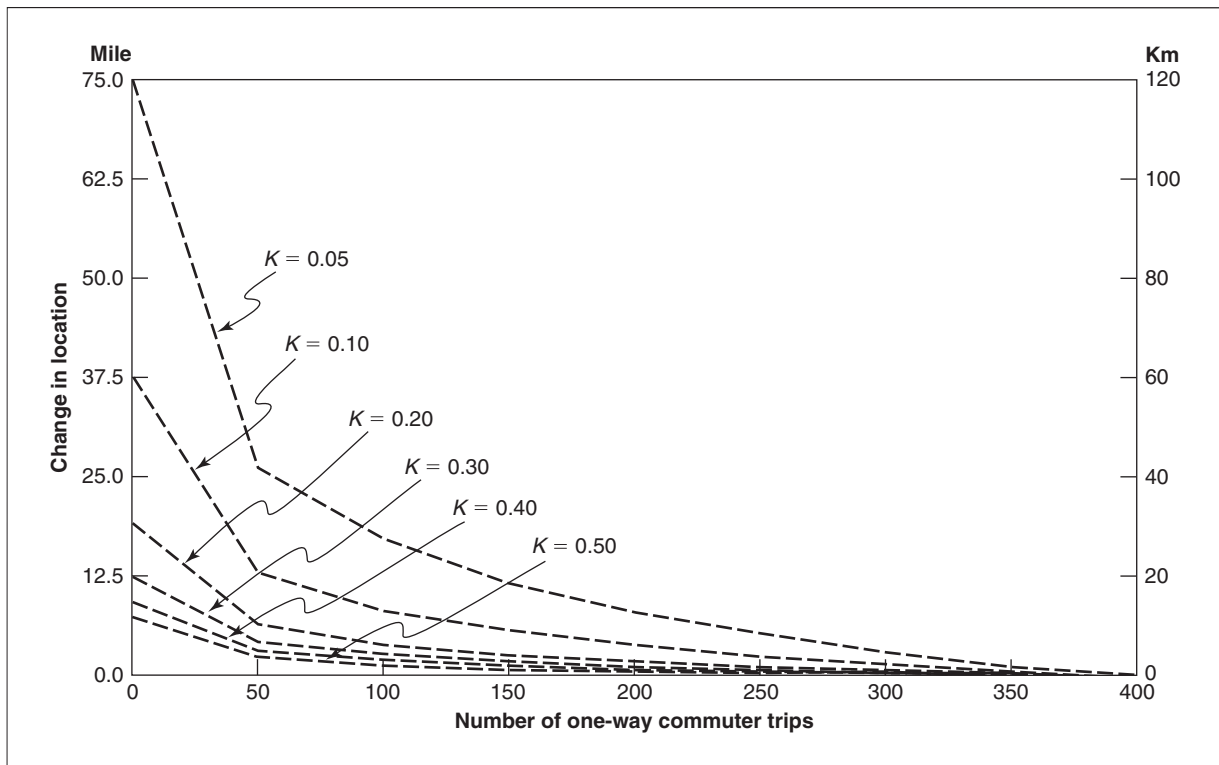
Notice that this relationship consists of a constant term that does not vary with commuting trips per year, minus a term that increases logarithmically with the number of annual commuting trips.

How would residential location change with the onset of telecommuting? To examine this, we define the change in least-cost location,

$$\Delta d^* = d^*(V_1) - d^*(V_0)$$

Replacing Equation 2.29 into this definition yields

$$\Delta d^* = [\ln V_0 - \ln V_1]/K = \ln (V_0/V_1)/K$$

Figure 2.13 CHANGE IN RESIDENTIAL LOCATION WITH TELECOMMUTING

SOURCE: Lund and Mokhtarian (1994). Reprinted with permission.

Note that this change in equilibrium location is affected by only the change in commuting trips and the decay constant of land prices. Other factors entering into the initial location decision do not affect the magnitude of change in the equilibrium least-cost location (Bonsall and Shires 2006).

Example

Consider a household initially located 6.25 miles (10 km) from the metropolitan center ($d_0^* = 6.25$ mi) where 400 one-way commuting trips are made annually ($V_0 = 400$). Land prices decay exponentially at a constant rate ranging from 8 percent to 80 percent per mile (5 percent to 50 percent per km) or $K = 0.08$ to 0.8 per mi. Figure 2.13 shows the change in equilibrium residential location as a function of the number of commuting trips and land prices. It confirms the theoretical and intuitively appealing finding in Equation 2.27, that residential location is affected most by telecommuting in a sprawling city with long commuting distances. ■

V. THE LOCATION DECISION

The above residential location discussions, particularly Equation 2.26, can be carried over to industrial activities.⁵ Assume that all activity takes place on a featureless plain consisting of land of equal quality. The rent that any producer will be prepared to pay for a given unit of land i , r^i , will be determined by its output

(the number of customer visitations) V , the price per unit at the market, γ , direct cost of production, c , the transport rate per unit of distance a' , and d_i , distance from the market:

$$r^i = V(\gamma - c) - Va'd_i \quad (2.30)$$

Here V , γ , c , and a' are assumed constant under conditions of perfect competition. This maximum rent, also referred to as bid-rent by Alonso (1960), is determined uniquely by the location of the site.

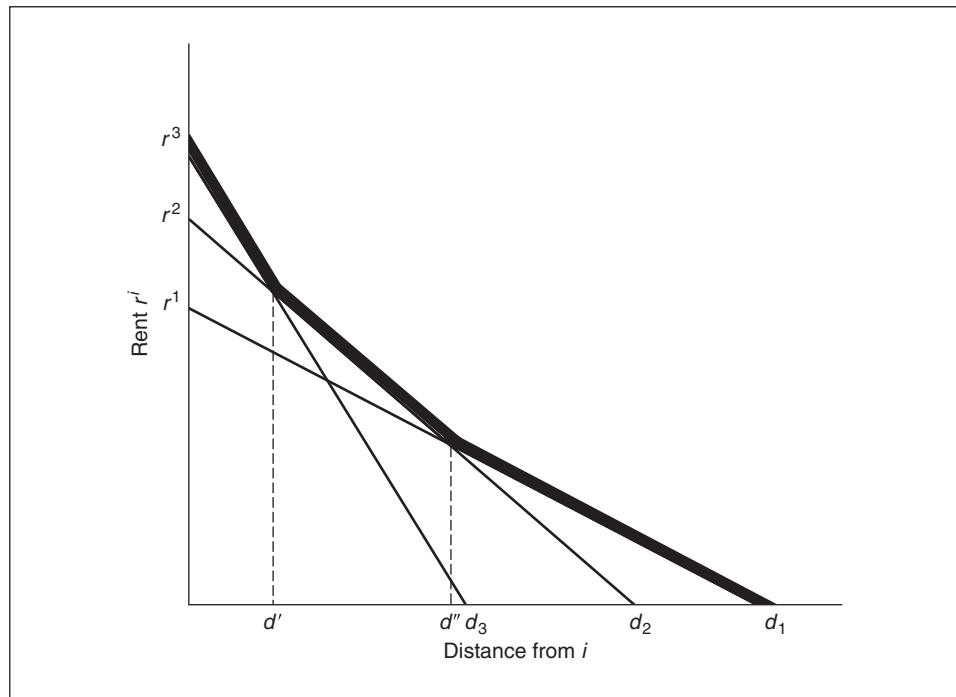
A. Bid-Rent Curves

Thus far we have assumed a single activity. If we introduce a second activity, it is obvious that V , γ , and c will not be constant and also it is likely that a' will vary according to weight or any special carriage requirements of the product. However, since perfect competition and freedom of entry prevail, we would not expect the profitability at the most favored location, which we can assume to be arbitrarily close to zero, to differ. The reason is that it and all producers would change production with consequent changes in price to restore an equality of profit. Hence the only change to be made if we have more than one activity is to introduce a' , the transport rate, as a determinant of r^i . It is then obvious that by knowing the transport rates for commodities we can derive the location pattern of production about the market. High transport cost activities will locate at a close distance and low transport cost activities will take locations further away. We can determine a relationship between r and d for each a' ; the maximum r^i payable at each d_i will determine the activity which will locate there.

Following Alonso (1964), this is best illustrated with a series of bid-rent curves as shown in Figure 2.14. Each bid rent curve $r^i d_i$ is defined by the linear Equation 2.30. Points d' and d'' define important switch points in land use between activities with different bid-rents. The piecewise linear line highlighted in bold is the revealed rent function for the area on the basis that land is allocated to the highest bidder.

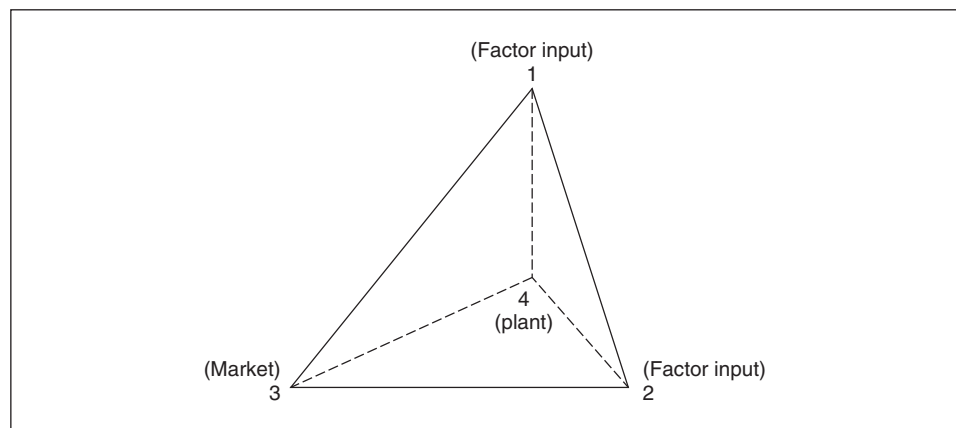
B. Industrial Location

Weber (Friedrich 1929) also started with the basic premise that particular locations do not have cost advantages in the actual manufacture of goods. However, in addition to land, most manufacturing industry requires inputs of more than one factor of production and, unlike land, these other factors cannot be assumed to be uniformly distributed in general. The location of a plant will therefore depend on the relative pulls of the various material locations and the market. Weber assumes these to be points rather than areas for simplicity. Assuming that for a particular product these various points are not coincident, the critical factors to be considered will be the relative weights of inputs and outputs and the distances over which these relative weights of input and outputs must be moved. Since transport rates depend on these two factors, the main interest was whether industries would locate nearer the market or to the source of materials and this could be related, through the transport costs, to whether the production process was weight losing or weight gaining. The materials index, the ratio of material

Figure 2.14 BID-RENT CURVES

weight to product weight, is a crude measure. It suggests that high values would involve a location dominated by sources of materials and low values (less than unity) would involve market domination, while values of about one would suggest location indifference.

The basic location criterion is thus minimizing total transport costs, assuming that market price of the product and prices of factor inputs are given and independent of location. The optimal location involves finding a set of

Figure 2.15 WEBER'S INDUSTRIAL LOCATION MODEL

distances d^i the inputs must be moved and distance-to-the-market D : $w_1 d_1 + w_2 d_2 + \dots + w_n d_n + D$. Here w_1 and w_2 , are the inputs required per unit of output. Figure 2.15 illustrates the simplest case of such a model. The figure shows a location triangle relating the market, node 3, to the two factor inputs at 1 and 2. The distances 3-1, 3-2, and 1-2 are geographic distances between the points. The optimal location for a plant at node 4 depends on the effective forces represented by the lines linking it to each corner. These forces are proportional to the relative weights of inputs or outputs as taken into account in the materials index. Node 4 can be found by constructing circles representing isocost lines centered on each corner of the triangle and examining their intersections. The most interesting result from this model is the dominance of end-points, many of which appear optimal, in-between points are of little importance. Numerical examples of this result are shown in Chapter 4.

C. Residential Location Models

According to Alonso (1964), the consumer looking for a housing location maximizes a utility function $v = v(x, s', d)$ where x is the quantity of a composite consumption good representing other activities engaged in by the consumer, s' is the average-size of site, and d is again the distance from the subarea of interest. In his/her location decision, the consumer is constrained by his/her available budget b^U , $p''x + r^i s'_i + a'd_i \leq b^U$, where p'' is the price of the composite consumption good. It is from this model that the bid-rent function for each individual can be derived as the maximum amount a person is willing to pay for a site that would be just as desirable as another.

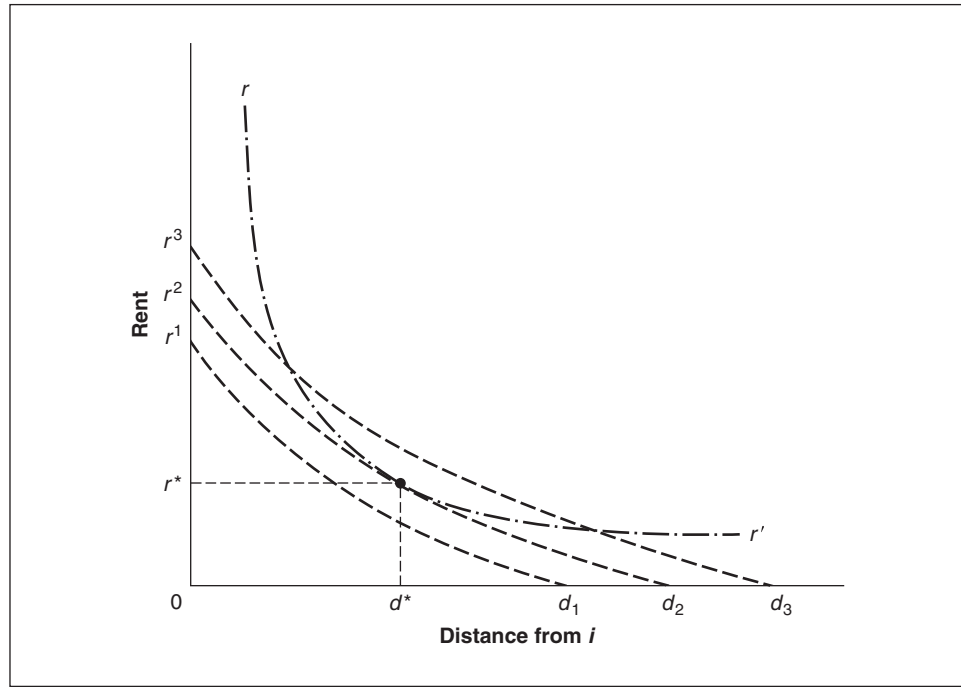
If we interpret the value of r^i in the above model as being the bid-rent for that location, then from the maximization exercise, we derive

$$\frac{\partial r}{\partial d} = \frac{p''}{s'} \frac{U_d}{U_x} - \frac{1}{s} \frac{\partial(a'd)}{\partial d} \quad (2.31)$$

where U_d and U_x are the appropriate marginal utilities of location and the composite consumption good. Rearranging Equation 2.31 in terms of marginal rates of substitution, we obtain

$$\frac{U_d}{U_x} = \frac{1}{p''} \left[s' \frac{\partial r}{\partial d} + \frac{\partial(a'd)}{\partial d} \right]$$

The above equation states the following: The incremental satisfaction from relocation (in terms of movement outward), which is obtained by substituting travel for goods, must be exactly equal to the cost of that relocation in terms of changing rent costs and changing travel costs. For simplicity we can assume that the good x has a price of unity such that $1/p'' = 1$. Furthermore, since the marginal rate of substitution is assumed to be conventionally negative and since transport costs will increase with distance, the land costs term must be negative. Obviously sites must always have a non-negative size and hence $\partial r / \partial d < 0$; we thus have the basic result that rents must decline with distance and hence the normal assumed shape of the bid-rent curve of Figure 2.16. In this figure, the lines $r^i - d_i$ represent bid-rent curves for an individual household. The higher the

Figure 2.16 HOUSEHOLD LOCATION MODEL

curve, the lower the level of satisfaction. The curve $r-r'$ is the equilibrium rent function for the city formed as an envelope curve to the various bid-rent lines of Figure 2.14. The equilibrium rent and location for this household is represented by (d^*, r^*) .

VI. SCALE AND NUMBER OF PUBLIC FACILITIES

Consider a homogeneous service to be distributed over some spatially distributed population. Let us assume that the service is distributed from a point-representable system of approximately up to four facilities— p_1, p_2, p_3 or p_4 —each having an identical scale \bar{P} measured in terms of capacity, capital outlay, or some other metric. The service is consumed by individuals who travel to the facilities for this purpose, and the service is priced at zero, meaning a public service provided by government to the citizens in the area. Total consumption Q of the service is the measure of effectiveness.

A. Static Short-Run Equilibrium

Now total consumption Q is a function of scale P and the number of facilities p

$$Q = Q(\bar{P}, p) \quad (2.32)$$

Total cost of the system C_t is made up of capital cost C_s and operating cost C_o , $C_t = C_s + C_o$, where

$$C_s = C_s(\bar{P}, p) \quad (2.33)$$

and

$$C_o = C_o(V) \quad (2.34)$$

In other words, capital cost depends on the number and scale of facilities built, while operating cost is related to the number of consumers served (V). The spatial pattern of facilities for a given (\bar{P}, p) is that pattern for which V is maximized. There exists a fixed budget b^U between capital and operating expenditures.

Figure 2.17, Figure 2.18, and Figure 2.19 illustrate some likely properties of Equations 2.32 through 2.34. Since the service is zero-priced, there is presumably some upper limit V^* to the amount that a population might be expected to consume. Holding the number of facilities constant in Figure 2.17, positive variation in scale may be expected to produce first increasing then decreasing positive variations in demand. The curves in Figure 2.17 actually represent a family of sections through the surface of Equation 2.32. They are therefore demand or consumer coverage curves for the service, given a fixed number, p_k , of facilities at varying scales. Scale expenditures play a role of negative prices or subsidies. An exactly analogous diagram could be made for the number of facilities, holding scale constant. The general character of $V(\bar{P}, p)$ is thus a function monotonically increasing to some asymptote V^* . It would look like a curved surface climbing away from the origin.

Figure 2.17 COVERAGE OF CONSUMERS

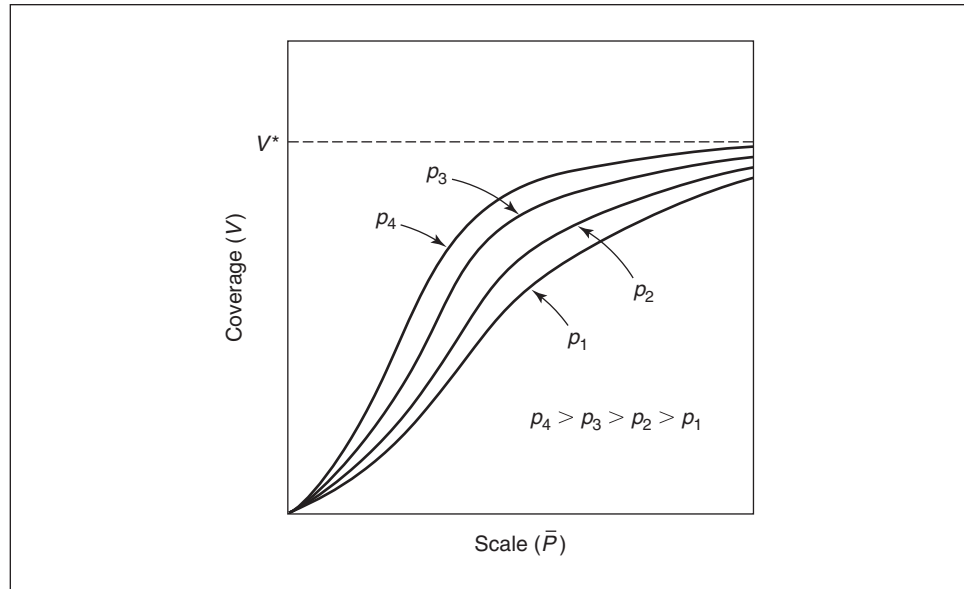
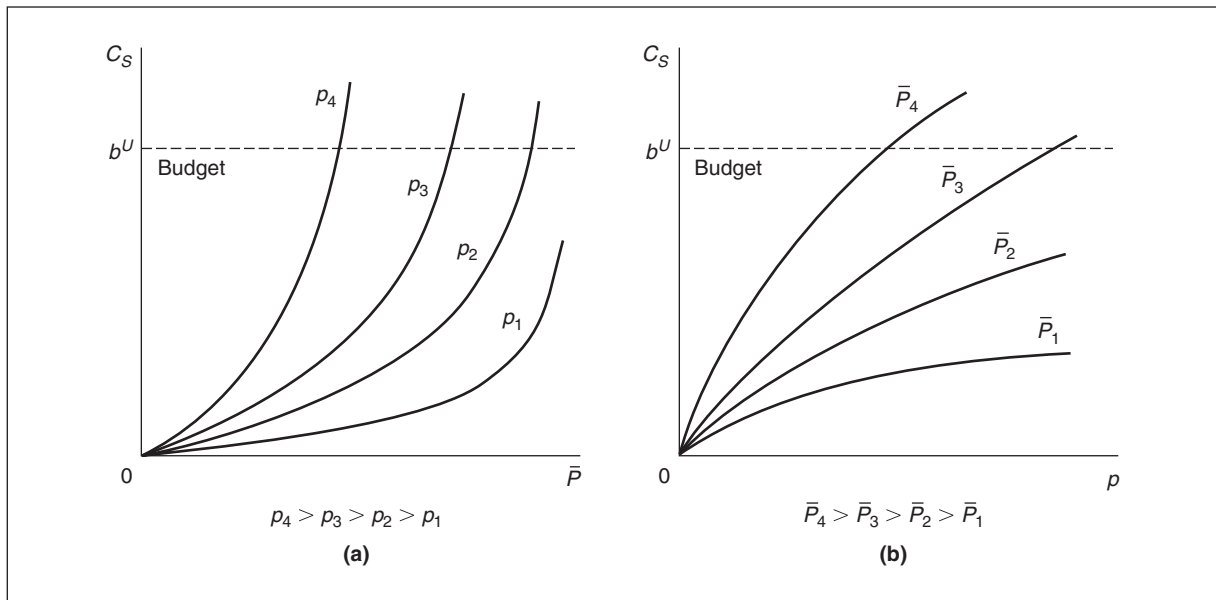
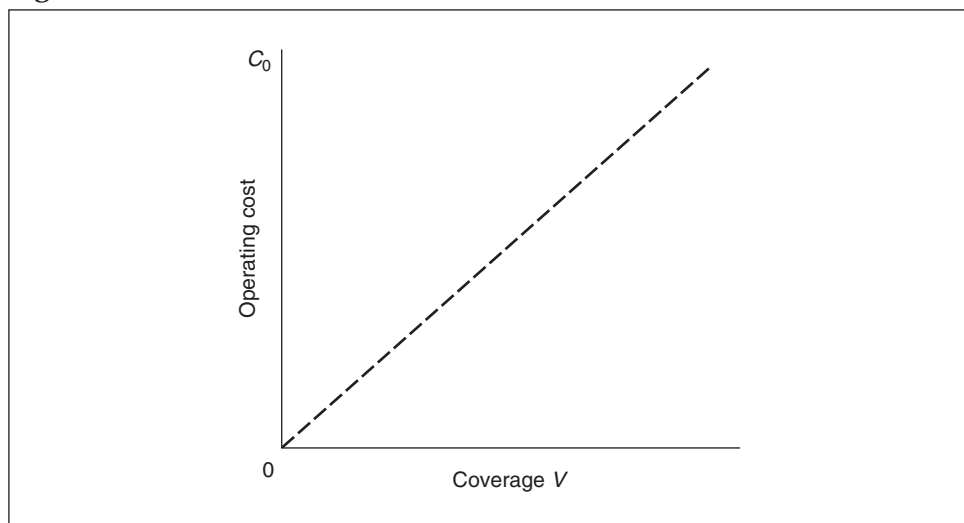


Figure 2.18 CAPITAL COSTS

Cost relationships may be handled in a similar way. Figure 2.18 presents a pattern of capital cost variations for constant levels-of-scale and number of facilities respectively. Although we assume that increase in scale eventually incurs higher marginal cost, there seems to be no reason for such an increase with the replication of facilities. Rather, the reverse seems to hold. The capital cost surface, C_s , may be generated from the families of sections in Figure 2.18. In short, increase in scale results in lower marginal cost compared with construction of new facilities in the beginning, and reverses itself as the system expands to full size.

Figure 2.19 OPERATING COST

For operating cost, C_o , several problems arise. We have made it a function of total demand on the assumption that the marginal product for any variable input to a given system does not vary with the form of the system itself, but only with the aggregate quantity of services demanded and produced. The reason for the distinction between capital and operating costs should be clear. The latter depends upon demand, representing the variable cost of responding to demand at the level induced by the former. In part, this may be an artificial distinction. Demand for a service does respond to the level of variable inputs insofar as it determines convenience and quality of service. We will avoid this complication for the moment by assuming that variable inputs are added to maintain some constant level of quality. For simplicity, this relationship is represented as generally linear in Figure 2.18, although it should be noted that in terms of the variables of Figures 2.17 and 2.18, it is likely to be nonlinear.

With appropriate assumptions about continuity and well-behaved functions, the problem may now be formulated as a constrained maximization: $\text{Max } V = V(\bar{P}, p)$ subject to $C_t = b^U$. The Lagrangian for this problem is $z = V(\bar{P}, p) - \lambda[C_s(\bar{P}, p) + C_o(V) - b^U]$ for which the conditions for maximization become:

$$\frac{\partial z}{\partial \bar{P}} = \frac{\partial V}{\partial \bar{P}} - \lambda \left[\frac{\partial C_s}{\partial \bar{P}} + \frac{\partial C_o}{\partial V} \frac{\partial V}{\partial \bar{P}} \right] = 0$$

$$\text{or} \quad \frac{\partial V}{\partial \bar{P}} = \left[\lambda / \left(1 - \lambda \frac{\partial C_o}{\partial V} \right) \right] \frac{\partial C_s}{\partial \bar{P}} \quad (2.35)$$

Similarly

$$\frac{\partial V}{\partial p} = \left[\lambda / \left(1 - \lambda \frac{\partial C_o}{\partial V} \right) \right] \frac{\partial C_s}{\partial p} \quad (2.36)$$

and

$$\frac{\partial z}{\partial \lambda} = C_s(\bar{P}, p) + C_o(V) - b^U \quad (2.37)$$

Combining Equations 2.35 and 2.36, we obtain the maximization condition

$$\frac{\partial V}{\partial \bar{P}} / \frac{\partial V}{\partial p} = \frac{\partial C_s}{\partial \bar{P}} / \frac{\partial C_s}{\partial p} \quad (2.38)$$

The equilibrium condition basically says that the maximal coverage is attained by a combination of scale expansion and new facility construction as justifiable by the marginal costs of the two ways to provide capacity. The consequences of our assumption about variable operating cost show up immediately in Equation 2.38. The equilibrium condition for demand maximization includes only system variables. If this seems peculiar, we might reflect that operating cost appears in Equation 2.37, which says that the cost for service coverage and system

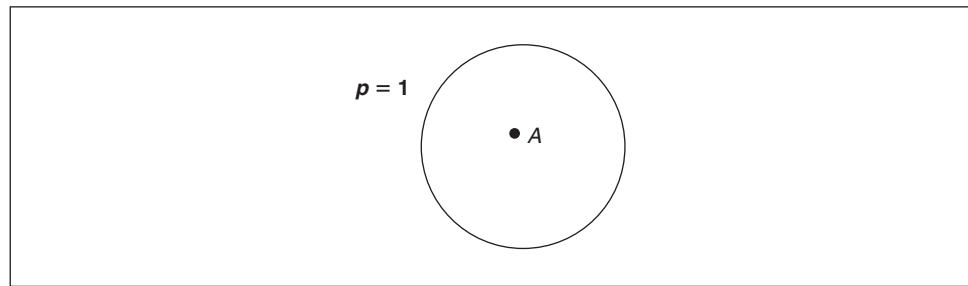
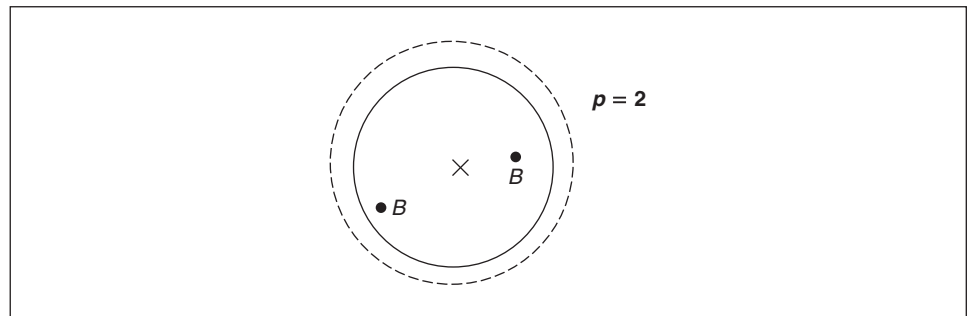
capacity expansion is limited by the budget available. Given our assumption that a given increase in demand generates the same operating cost no matter whether it derives from the scale or number of system components, its absence from Equation 2.38 is less surprising. Whether that assumption is tenable is another matter.

More significantly, this formulation evades the problem of location via its cost structure, which is totally dependent upon scale and number of facilities and has no spatial cost components. So far the researchers have been unable to incorporate the location problem into a pure analytic model. In view of the numerous mathematical programming and heuristic approaches to this type of problem, there would seem to be advantages to structuring the total problem as a computer model. In analytical terms, this raises the problem of our assumption of continuity in the variable p . Using a calculus-based model, we cannot simultaneously assume it would be continuous for scale analysis and discrete for a location-effective algorithm. Perhaps an iterative estimation process is the way around this problem, but the theoretical result is less precise. In any case, it seems probable that the location problem for public facility systems must be attacked in tandem with system structure and scale. Several problems still remain. Introduction of variable facility scales in a single system is clearly necessary. As soon as this is done, then questions of hierarchy begin to arise.

The static equilibrium treated above is general in the sense that it deals with simultaneous location and scale of all components of a facility system. The equivalent partial problem might be formulated in several ways. If an increment to a budget for an existing system is given, then we might be interested in determining the optimal addition to the system. This does not necessarily mean that any new components are added. The entire budget increment could be spent on scale changes. If the problem is to achieve a specified incremental gain in some effectiveness measure, the same qualifications would apply. In these circumstances it is not clear how a partial form should be specified. Possibly, it should hold the present facility location structure constant and allow only scale changes and new facility locations. Again, advances in more sophisticated methods than simple calculus are necessary for addressing such problems.

B. Dynamic Long-Run Equilibrium

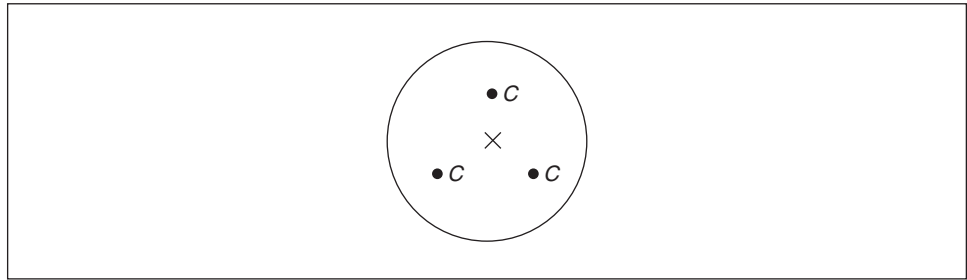
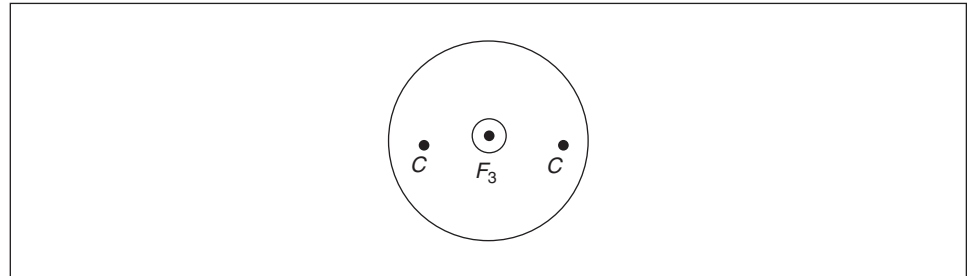
To analyze systems of facilities with static equilibrium analysis is to ignore a most important characteristic: their changes over time. Facility systems are usually built quite slowly, reacting to changes both in the size of the broader systems they serve and in technology and social preferences. If the broader system is a growing city, then there may be conflict between static and dynamic system optima. This may be especially true if, for whatever reason, decisions early in a system's development can effectively close off options for later forms. A geometric illustration of a dynamic system conflicting with static solutions is offered by a simple model. Consider the circular and generally symmetric city represented in Figure 2.20. At this particular size and for some local service, the optimal number of facilities is one, and it is located at the center, A . The city grows symmetrically both in density and at its outer margin until it reaches the size shown in Figure 2.21. At this new level the static-equilibrium solution, taking into account a probable larger budget for the service, calls for two identical facilities, B . If they

Figure 2.20 FACILITY EXPANSION IN A CIRCULAR AND SYMMETRIC CITY**Figure 2.21** LARGER CITY WITH TWO FACILITIES

are located symmetrically, there is no path of growth for this facility system from stage 1 to stage 2 that does not call for removal of *A*. Whether that is likely depends on the rate of growth and the fixed capital investment in *A*.

The example is made artificial by the assertion of identical facilities. In practice, accommodation may be partially achieved by variations in scale among facilities. For example, the equivalent problem for three components might be to approximate a symmetric uniform scale optimum by a variable scale but still symmetric three-component system (see Figures 2.22 and 2.23 respectively). In the latter, the original facility is retained at a larger scale than the others. Without specifying particular forms for the relationships between spatial pattern, scale, and demand, we cannot say much more than this.

The dynamic long-run equilibrium discussion above suggests two modeling approaches. We may look for possible system growth paths through time under varying constraints and criteria for effectiveness and try to identify stages at which such paths coincide with static equilibrium solutions, or we may set up static equilibrium solutions and try to construct minimum cost paths to connect them. Since most facility system analyses are likely to start with an existing set of components, most of which incur high relocation costs, either form could be employed. The choice is perhaps yet another version of the process/end-state conflict in planning models, in this case with both forms involving specific criteria for choice since the decisions are public. Very little work in this direction has been done. Chan (2005) discusses growth paths of land use, rather than facility location, in his chapter on "Bifurcation and Disaggregation." The continuous generalization of facility location—land use—is easier to model inasmuch as it avoids the discreteness or lumpiness

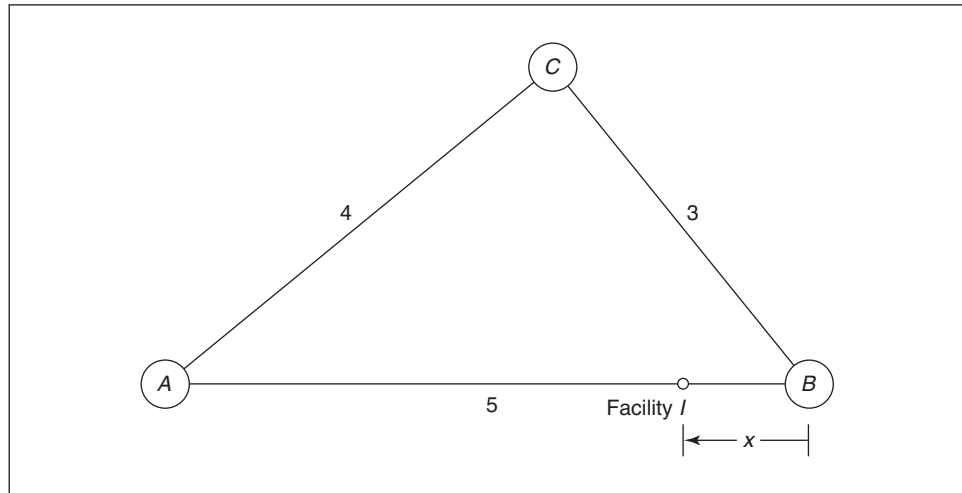
Figure 2.22 THREE FACILITIES AT UNIFORM SCALE**Figure 2.23** THREE FACILITIES AT VARIABLE SCALE

that prevents smooth transition from stage to stage, although bifurcation models do allow for precipitous happenings to take place. Again computer models seem to be most promising given the mathematical complexity of any reasonable looking structure for analysis. (One such program, the Garin-Lowry model, is included on the CD/DVD under the YI-CHAN folder.)

The main problem of locating public services, as can be seen, is choosing the scale and the number of facilities at specified geographic locations that would be most adequate to provide the public services for the budget allocation. The theoretical exposé, while addressing most of the key considerations in planning for public services, has to be further refined for specific applications. Associated with the scale and location considerations, for example, are the ways and means to make the public service available to the community. In this regard, the spatial location of a facility becomes as important as the scale and the number of facilities.

VII. SPATIAL LOCATION OF A FACILITY

Consider the triangular network ABC as shown in Figure 2.24, where there are three highways represented by the three edges of the triangle. A facility, for instance, a shopping mall, is to be located on the highway system so that the distance to the farthest population center A , B , or C is minimized. The demand at A , B , or C does not enter into the picture in this example; only distances are considered.

Figure 2.24 TRIANGULAR NETWORK *ABC*

A. Center of a Network

Suppose for the time being the facility is to be located among candidate sites on a highway between nodes *A* and *B*, which has a separation of 5 miles (8 km). Let us place a facility at point *I* at a distance of x from node *B*. The distance function between node *A* and point *I* is $5 - x$, and the distance function between node *B* and *I* is simply x . These distance functions are shown in Figure 2.25 (Ahituv and Berman 1988). We are supposed to find the one center location, or the location which minimizes the farthest point away. The maximum distance is shown on the upper envelope of Figure 2.25. The minimum occurs at $x = 2.5$ miles (4 km) from *A*, or halfway between *A* and *B*, which is located at the lowest point on the envelope. This is sometimes referred to as the **mini-max** solution.

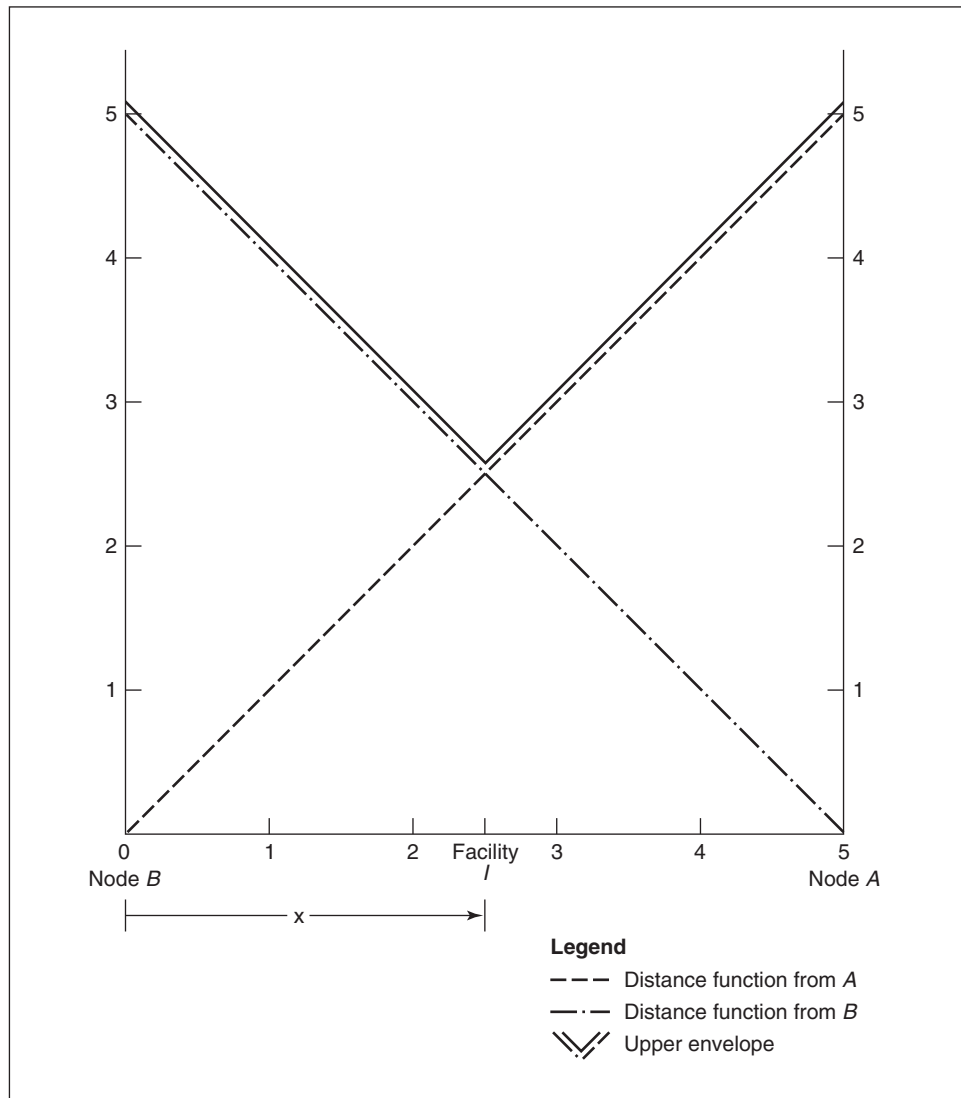
Unfortunately, the problem is more involved, since there is node *C* as well. Let us examine the distance between points on link (*A*, *B*) and node *C*. If the facility is located at node *B*, the shortest distance to node *C* would be 3 miles (4.8 km). When we move point *I* along the link (*A*, *B*) from *B* toward *A*, the shortest distance function becomes $3 + x$. This, however, stops when x reaches 3 miles from node *B*, because at that point it is better to approach node *C* via node *A*. The distance function from *I* to *C* becomes $9 - x$, where 9 is the sum of the distances of links (*B*, *A*) and (*A*, *C*), and x remains to be the distance of point *I* from node *B*. The complete distance function is given by

$$d_{I3} = \begin{cases} 3 + x & \text{for } 0 \leq x \leq 3 \\ 9 - x & \text{for } 3 \leq x \leq 5 \end{cases} \quad (2.39)$$

The function is shown in Figure 2.26.

In Figure 2.27, we have combined the distance functions to nodes *A* and *B* from Figure 2.25 with the distance function to node *C* in Figure 2.26. A new upper envelope is drawn, which describes the maximum distance from *I* to nodes *A*, *B*, and *C*, depending on the location of *I* on link (*A*, *B*). The minimum of the maximum distance is obtained when the facility is placed at a distance of $x = 1$ mile (1.6 km) from *B*. At this facility location, the maximum distance to demands at *A*, *B*, and *C* is minimized at a value of 4 miles (6.4 km). In a similar

Figure 2.25 DISTANCE FUNCTIONS BETWEEN A FACILITY AND DEMANDS AT A AND B



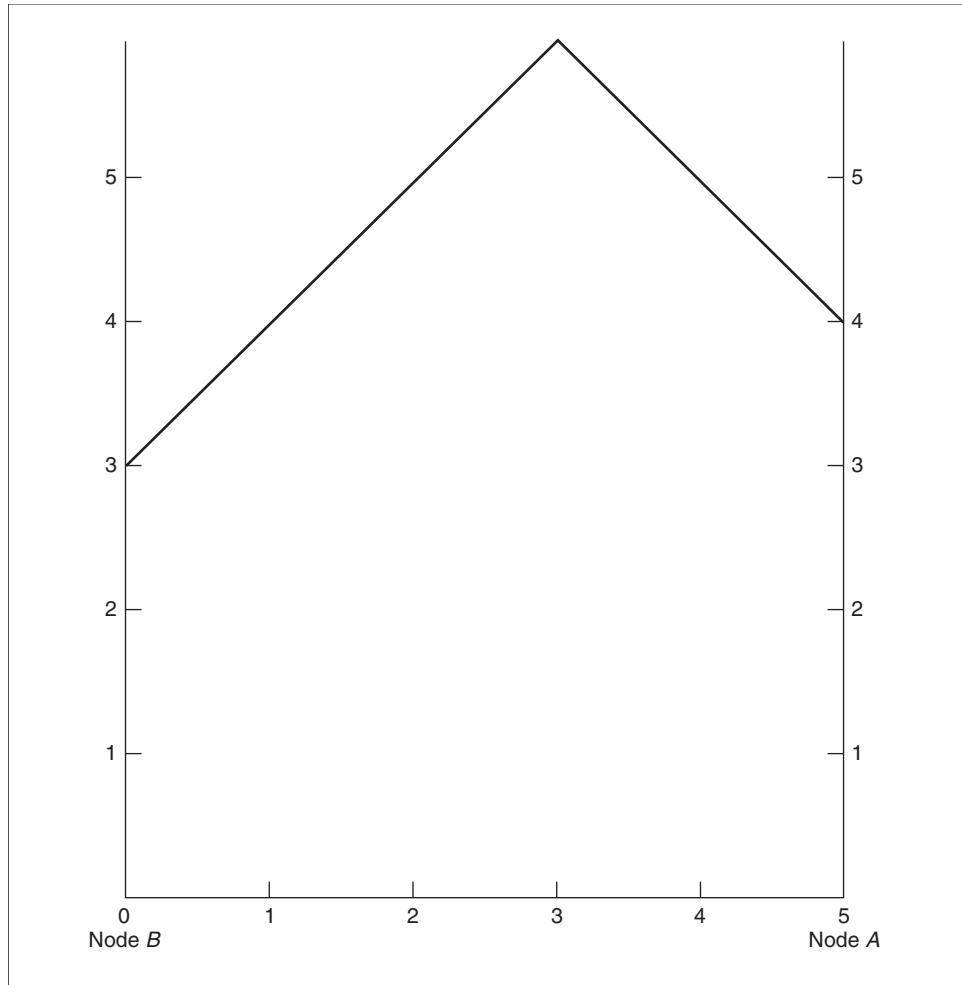
SOURCE: Adapted from Ahituv and Berman (1988). Reprinted with permission.

fashion, we proceed to inquire about the distance functions between points of link (B, C) and node A, then link (C, A) and node B. The process is in fact quite tedious. More efficient algorithms are available to circumvent this exhaustive search procedure, but they are beyond the scope of this text. Interested readers are referred to the “Facility Location” chapter in Chan (2005).

B. Median of a Network

Suppose we are to locate a facility such that the average distance from a demand node to the nearest facility is minimized—the minimum-of-the-weighted-sum (**mini-sum**) solution. It has been shown (Hakimi 1964) that such a facility has to

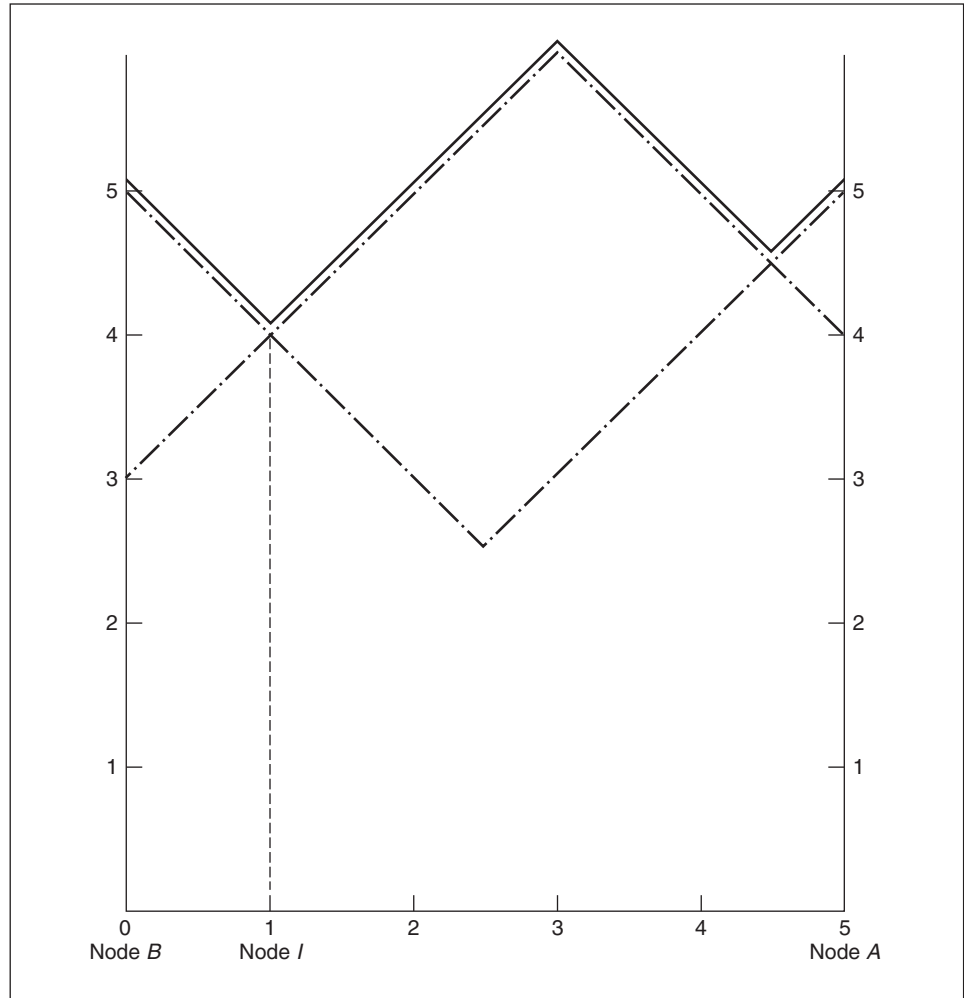
Figure 2.26 CENTER DISTANCE FUNCTION FOR LOCATING FACILITY IN A NETWORK



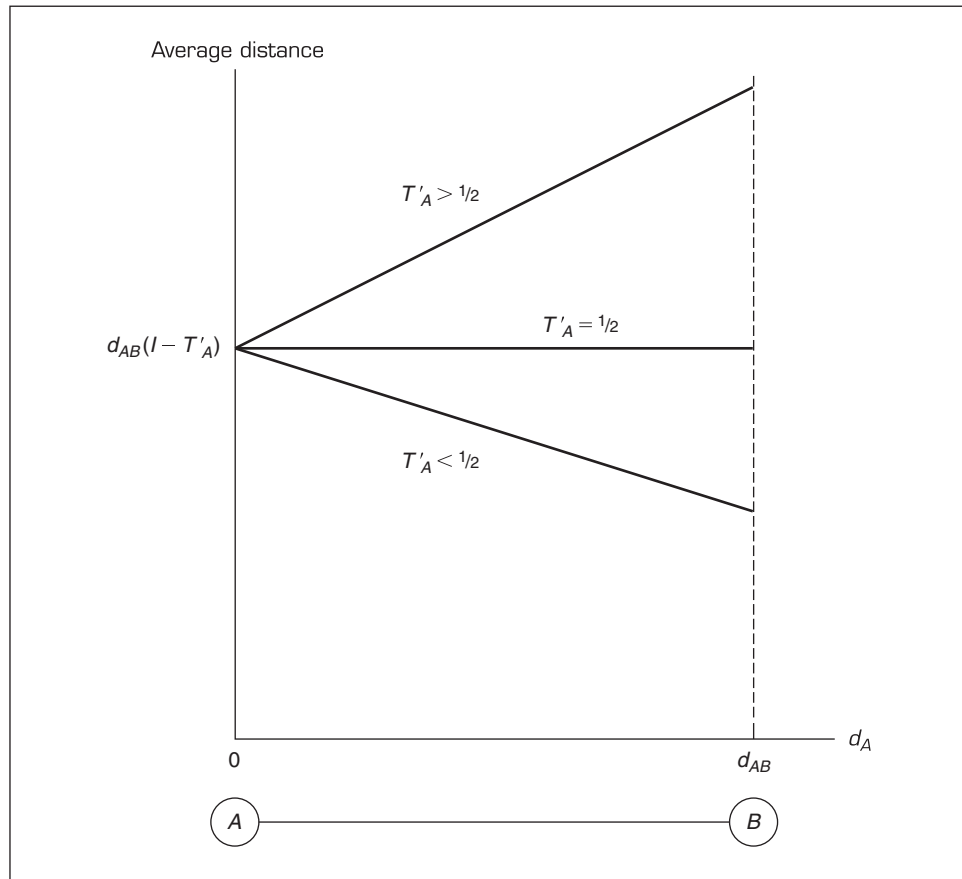
SOURCE: Adapted from Ahituv and Berman (1988). Reprinted with permission.

be located at a node. This is distinctly different from the center problem above, in which the facility can be anywhere on an arc (including the two nodes that define the arc also.) To show this nodal optimality condition for one median, we examine the network consisting of only one link, as depicted in Figure 2.26 (Ahituv and Berman 1988). A and B represent the two demand nodes, which are separated by a distance d_{AB} . The demand proportion generated at node A is T'_A , while that at node B is $T'_B = 1 - T'_A$. Suppose we place the facility at I on link (A, B) . Assume d_A is the distance between node A and the facility I . The average weighted distance for delivering the service from I to the consumers, or for the consumers to access the facility, is

$$T'_A d_A + (1 - T'_A)(d_{AB} - d_A) = T'_A d_A - d_{AB} - d_A - T'_A d_{AB} + T'_A d_A = d_{AB}(1 - T'_A) + d_A(2T'_A - 1) \quad (2.40)$$

Figure 2.27 COMBINED DISTANCE FUNCTION FOR FACILITY IN A NETWORK

The first term of the above equation is constant; it does not depend on the location of I . The second term is a function of the location of I , or d_A . Now suppose node A generated more demand than node B , thus $T'_A > 1/2$. Hence $(2T'_A - 1) > 0$ and Equation 2.40 is minimized when $d_A = 0$, or when the facility is located at A . However, if node B generated more demand than A , namely $T'_A < 1/2$ and $(2T'_A - 1) < 0$, Equation 2.40 is minimized when d_A assumes its biggest possible value d_{AB} . In this case we will place the facility at B . If the two nodes generate equal demand, facility I may be located anywhere on link (A, B) including the two nodes. Figure 2.28 illustrates the above problem graphically. The average distance as represented by Equation 40 is plotted as a function of the distance from node A to the facility I , d_A . For $T'_A < 1/2$, the median should be located at A , where the average distance is minimized. For $T'_A = 1/2$, the median can be anywhere between A and B and the travel distance is the same. For $T'_A > 1/2$, the facility should be located at B . Figure 2.28 contrasts sharply with Figure 2.25 in that upper envelope in

Figure 2.28 AVERAGE DISTANCE AS A FUNCTION OF MEDIAN LOCATION

SOURCE: Ahituv and Berman (1988). Reprinted with permission.

the latter has a kink in the middle while the former is a monotonically increasing or nondecreasing function. The former identifies a nodal optimum at either A or B , while the latter locates an optimum in between the two nodes A and B .

This problem will be discussed again in Chapter 4, where the same problem will be formulated as a linear program, which yields the nodal optimality results directly from the properties of a linear program. From the gravity model, center and median discussions, it is quite clear that depending on the figure of merit for evaluation, a facility can be located at very different places. It is therefore important to properly define an evaluation measure from the beginning of an analysis.

C. Competitive Location and Games

Let us now illustrate competitive location decisions on a network. Suppose there are already p facilities located. We wish to locate r new facilities that are to compete with the existing facilities for providing service to the customers at the nodes. All demands are perfectly inelastic and the consumers' preferences are binary. We assume customers will change their habits and use the closest new facility if and only if it is closer to them than the closest old facility. Ties

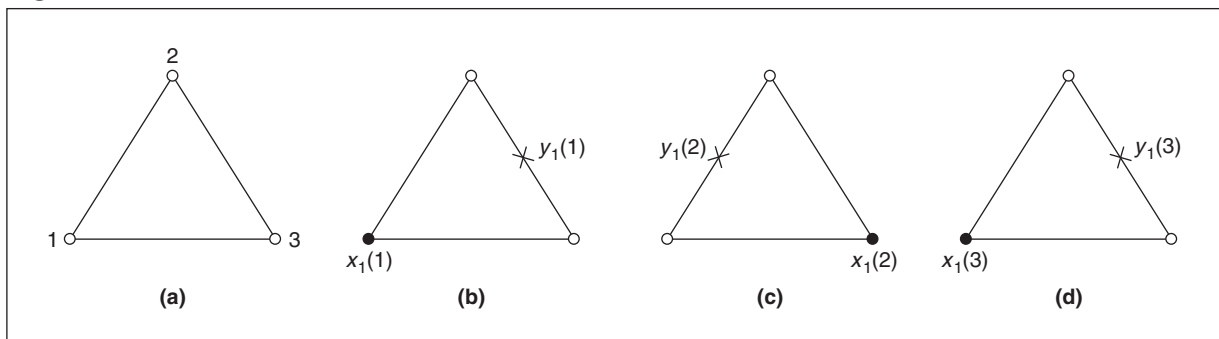
are broken in favor of an old facility. Suppose there are two competitors, where both players wish to control as large a share of the market as possible. The first player selects p points for his facilities; the second player, having knowledge of the competitor's decision, selects r points. As the problem is presently stated, each player has exactly one move and has to make the best move possible. This is especially true in situations where the facilities are expensive to construct, and once the facilities are constructed no further moves can be contemplated. The first player knows that once the p sites are selected, the second player will then select the best possible r sites for the facilities. One may pose two possible scenarios for this game to continue beyond the first move by each player (Hakimi 1990).

- (a) The facilities are mobile but for each player it takes a certain amount of time to respond to the other player's choice of sites (move), assuming that the players do have the computational power to make the best move at each step.
- (b) The first player does not have the computational power to find r centers while each player does have the capability of finding r medians or p medians. For both cases, the question arises about where the two players will end up.

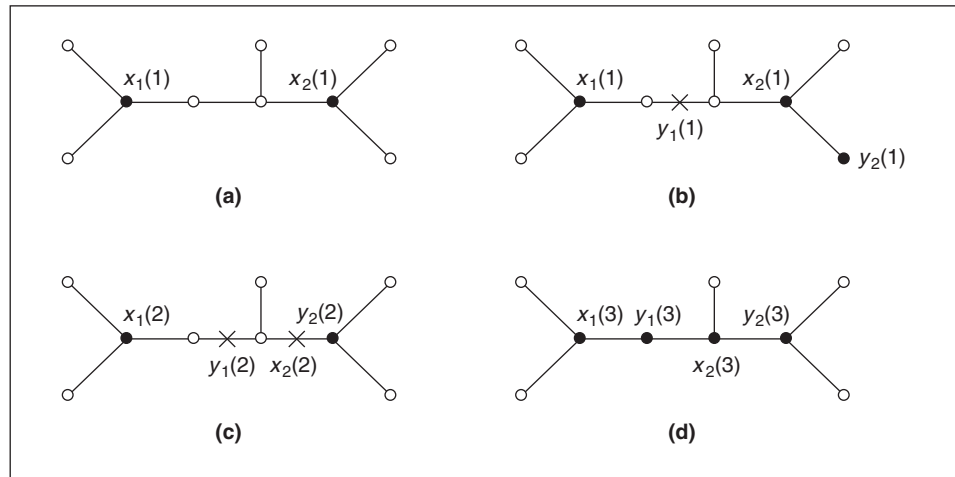
Example 1

In the example shown in Figure 2.29(a), we assume $p = r = 1$, the payoff at each node to be 1, and the arc lengths are all 1. In Figure 2.29(b) both players' first moves are indicated, where $y_1(1)$ is the mid-point on the edge (2, 3) which is a 1-median. At this stage, it is the first player's turn to move. That move ($x_1(2)$) and the second player's response to it ($y_1(2)$) are shown in Figure 2.29(c). Finally, Figure 2.29(d) indicates the third move of the first player and the second player's response. At this stage, it is clear that the game will continue indefinitely. Whichever player quits first is the loser and will control exactly one-third of the market, leaving the rest to the other player. This example illustrates a situation where the game does not reach an equilibrium, that is, where each player finds that continuing to move is the only way to avoid being limited to the one-third share of the market. Note that in the above example, the first move by the first player, that is the choice of $x_1(1)$, is a 1-center of the network. ■

Figure 2.29 NON-EQUILIBRIUM EXAMPLE



SOURCE: Hakimi (1990). Reprinted with permission.

Figure 2.30 EQUILIBRIUM EXAMPLE

SOURCE: Hakimi (1990). Reprinted with permission.

Example 2

Let us now consider the network of Figure 2.30(a). Assume the payoff at each node is 1, $p = r = 2$, and the arc lengths are all 1. The first player's move $\{x_1(1), x_2(1)\}$ and the second player's response $\{y_1(1), y_2(1)\}$ are shown in Figure 2.30(b). The first player's second move $\{x_1(2), x_2(2)\}$ and, correspondingly, the second player's second move $\{y_1(2), y_2(2)\}$ are shown in Figure 2.30(c). The players' third moves $\{x_1(3), x_2(3)\}$ and $\{y_1(3), y_2(3)\}$ are again shown in Figure 2.30(d). Now it is the first player's turn again. He or she knows, of course, of the positions of his or her competitor and finds that his or her present location is at a 2-median. Thus he or she will not move from his or her present position which implies that the second player also will not move and the game is over. Thus the game terminates in an equilibrium state. We note in passing that the first player's position constitutes a 2-center location of this tree network as well. ■

D. Imperfect Information

It can be seen that the spatial games illustrated above is based in part on the players' lack of perfect information. We start with a single player making a decision on the basis of a known set of information. The first decision to make is whether the player is in the best situation achievable. If he or she is not then he or she must take action. However, there are two problems: one is a lack of perfect information, such that what the player perceives is not necessarily true and because of this ignorance additional information might be needed. Secondly, the player recognizes that even if the adjustment to improve the situation is made, that might not be achieved in a given decision period. In general, our decision maker is assumed to be extremely myopic, to the extent that the system state does not change as a result of his or her decision. We have a situation reminiscent of early attempts to solve classic models of oligopoly markets, in other words, markets dominated by several players. Under these assumptions, the market solution could be shown to be stable, as in the Cournot case where the rival's output is assumed constant in each decision period, and one adjusts his or

her output to maximize profit accordingly. Or the market may be unstable, as in the Bertrand or Edgeworth case, where the rival's spatial prices are assumed constant, and the price is adjusted to undercut the rival.⁶ In these cases we need to examine two features, whether a full stable equilibrium will be reached and, if so, the speed at which this will take place. The critical factors will be the adjustment to the assumed optimal position and the possible error in making that assumption. Oftentimes, we are concerned less with the final equilibrium itself and more with the path leading to it. In this regard, we are particularly interested in individual players' reactions in each period (Vickerman 1980).

A more realistic model would need to relax the assumed myopia of individuals and introduce strategic reactions of the type adopted in game theory, in which perfect knowledge is assumed. Starting with pure zero-sum games, for example, a conservative player is maximizing minimum gain while the other equally conservative player is minimizing maximum loss. As implied in a zero-sum game, gain to one player matches the amount of loss to the other. In general, individuals are concerned not only with their own attempts to optimize but also with any reactions of conflicting parties to their own actions. A simple example will illustrate the complexities introduced here. A supermarket chain siting a new store will recognize that other shops will be responding to the same stimuli (for example, relative proximity to a new residential area) and that this may generate additional benefits such that the precise site cannot be planned independently. It also realizes that competitors will also respond in an attempt to secure new markets themselves. The calculation depends additionally on the assumptions made about the response of customers, both existing and potential. In the absence of collusion, all of these responses have to be given ahead of time, but the final solution will depend on how good those assumptions are. Once again we shall need to be concerned with whether the path converges ultimately to a stable equilibrium and the speed at which the adjustment takes place. In this case it is not sufficient simply to take assumed responses and examine the behavior of the system, since non-myopic individuals concerned with improving their situations will also learn from revealed responses and accordingly may modify their responses in subsequent decisions. Hence, we also require a learning process within the model.

It will be clear even from this simple description that a representative model of this type will be unavoidably complex. While it would be possible to proceed with continuous functions in a model, there is much to be said for taking a programming approach—an approach which involves systematic computational procedures (often using a computer.) Many of the decisions are of a discrete nature and may involve thresholds and discontinuities that are awkward for a continuous model. The use of discrete time periods also accommodates varying degrees of myopia in adjustment. It is also important that we should stress the operation of the economy as a series of explicitly individual but interdependent decisions. The most useful approach to this type of problem is **recursive programming**, in which a relationship between given system states and expected actions is established, and so are the attempts to simulate a sequence of expected actions through time (Nelson 1971).⁷

There are two possible assumptions about how the markets move into equilibrium at the end of each period. One way is to require the markets to clear period by period, so that a sequence of temporary equilibria is formed, or so that disequilibrium can exist. This was illustrated in Section VI of this chapter, where the transition between one, two and three facilities in a growth environment is anything but continual. An assumption of equilibrium appears unrealistic and

almost contrary to the logic of an adaptive model that depends on the independent, albeit linked, reactions of different individuals. Unrealistic as it may be, it does have a number of convenient, simplifying properties. For example, it raises the question of whether individuals attempt to move into full equilibrium. If experience teaches them to modify their behavior, it should also reveal the degree of success of such modification. Given these behavioral adaptations, a policy of suboptimizing may be less costly than an attempt at complete optimization. The sets of reactions might incorporate information about this learning process in a full disequilibrium, wherein it is a conscious decision of individuals that causes the failure to achieve market equilibrium.

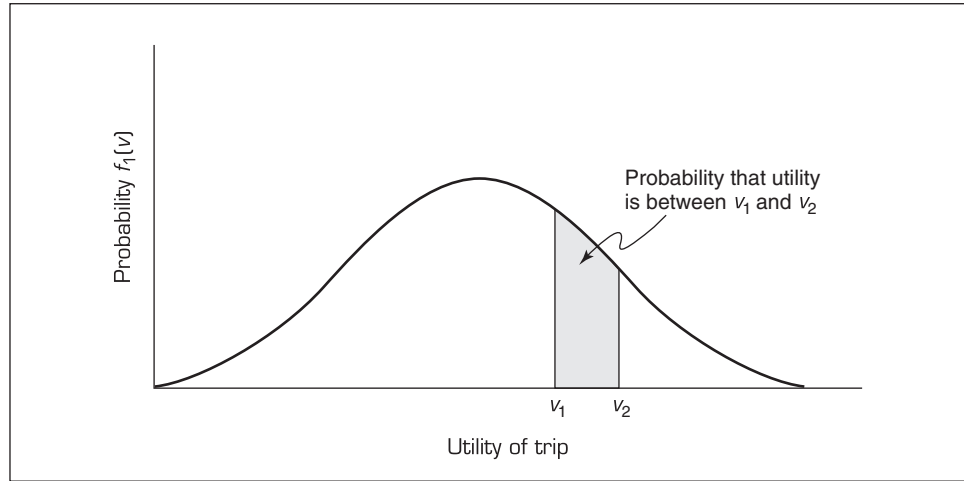
It will be apparent that this approach enables a considerable degree of flexibility in the structure and design of a model of the urban, and general spatial, system. At this level of generality it is not possible to draw even qualitative conclusions about whether the results will differ substantially from those of an equilibrium model. It does, however, seem reasonable to expect that, freed from a requirement of a dynamic equilibrium path or even a period by period establishment of equilibrium, the spatial economy may well exhibit a rather different structure. The next step is therefore to use simple versions of this model to simulate the development and structure of urban areas under, for example, different reaction schedules. Such an approach may form an empirical base in the examination of the performance and structure of urban economies under practical planning regimes. A further question is the extent to which such a model can be used to evaluate urban changes, given most evaluation procedures are based on equilibrium metrics. For further details, see chapters starting with "Generation, Competition and Distribution" and ending with "Spatial Equilibrium and Disequilibrium," in Chan (2005). (The reader may also wish to experiment with the software and data contained in the attached CD/DVD under the YI-CHAN folder.)

VIII. ECONOMIC BASIS OF THE GRAVITY-BASED SPATIAL ALLOCATION MODEL

In the traditional literature, the most common location technique for land use (as contrasted with facility location) is the gravity model. Here we will derive the various forms of the gravity model based on the assumption that individuals maximize their net benefits in choosing a destination facility (Cochrane 1975). The trade proportions among competing shopping centers, for example, reflect the overall probability of trips being made on the basis of the attractiveness and convenience of the shopping center. Various forms of gravity models have been proposed. They are reviewed below in preparation for later parts of this book.

A. The Singly Constrained Model

Singly constrained gravity model is one in which the number of trips originating in any subarea is assumed determined and fixed. These trips are being made to any of the competing facilities that offer the service. In addition, the model assumes that at each destination there exists some quantity of activities that attracts consumers to patronize that facility. Thus the activity at a shopping mall may be the size of the mall measured in retail floor space. We do not know the

Figure 2.31 PROBABILITY DENSITY FUNCTION OF TRIP UTILITY

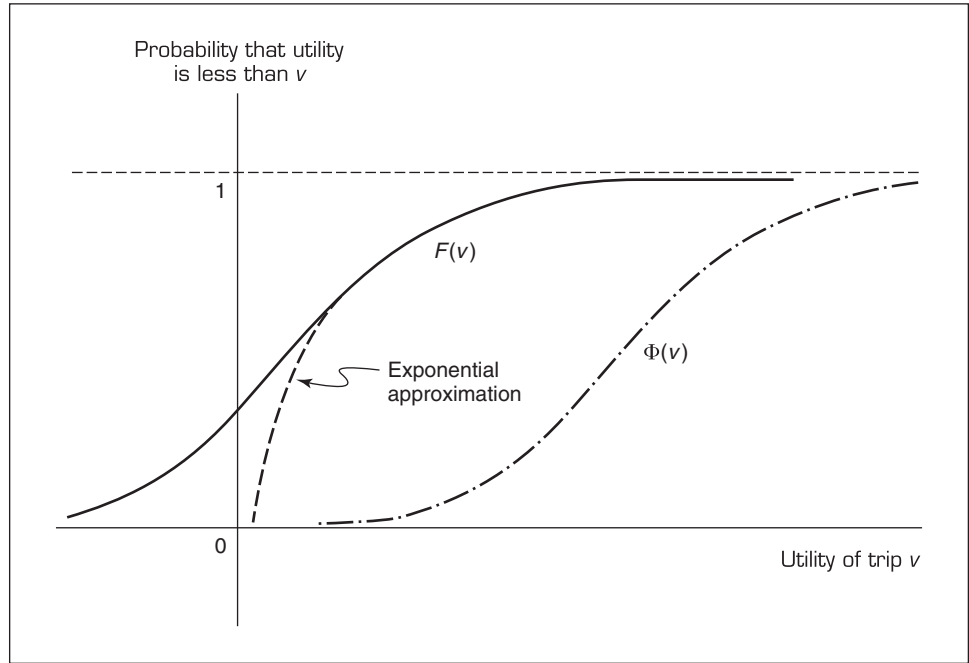
SOURCE: Cochrane (1975). Reprinted with permission.

precise value a trip maker might place on any particular trip, since tastes are individual. However, we hypothesize that we can assign a probability that this value will fall between trip utilities v_1 and v_2 (see Figure 2.31.) Define consumers' surplus as the net benefit of any trip after the trip cost has been subtracted from the basic value or utility. Since we can estimate the cost of any particular trip, we can estimate the probability that the surplus lies between any two values.

The central assumption of the present derivation of the gravity model is that the probability that a particular trip maker from one subarea will travel to a facility is the probability that the trip to that facility offers a surplus greater than that of a trip to any other facilities. The probability of an individual trip to a facility being optimal increases with the activity or opportunity at that facility and decreases with travel distance, since the net benefit is reduced by a greater cost. We consider the effect of the number of opportunities offered by a facility. Since we are interested in the probability that the trip to the facility is the best choice, we first estimate the probability of the utility of the optimal (highest utility) trip lying within particular bounds. The cumulative distribution function of the largest v among n independent samples from a common underlying distribution is given by $\Phi(v) = [F(v)]^n$ where $F(v)$ is the cumulative distribution function of the common underlying distribution. The reason is that the cumulative distribution function is the probability that the value is less than or equal to v , and the probability that the best of n is in this range is identical to that of all n being less than or equal to v . Now provided n is moderately large (in double figures at least), $\Phi(v)$ is scarcely affected by the shape of the underlying distribution outside the upper tail (see Figure 2.32.) It is possible to develop an asymptotic (large n) expression of $\Phi(v)$ based only on the shape of the upper tail. If the upper tail can be approximated by a simple exponential function, as indicated in Figure 2.32, $\Phi(v)$ rapidly approaches the simple asymptotic form

$$\Phi(v) = \exp[-ne^{-b(v - \bar{v})}] \quad (2.41)$$

where \bar{v} is the average trip utility.

Figure 2.32 CUMULATIVE DISTRIBUTION FUNCTIONS OF TRIP UTILITY

SOURCE: Cochrane (1975). Reprinted with permission.

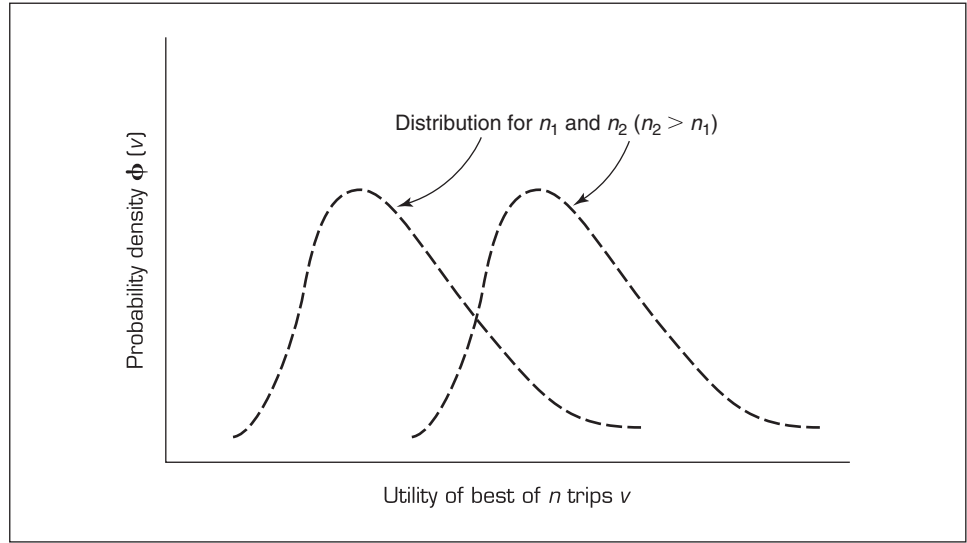
Provided that we assume only that the underlying distribution is approximately exponential in the upper tail, the probability density function for the utility of the best trip in any subarea is given by the differential of Equation 41. This distribution is indicated in Figure 2.33. It is a positively skewed distribution whose skewness is independent of b , \bar{v} and n . The mean is

$$\bar{v} + \frac{1}{b} [\ln(n) + 0.577]$$

and the standard deviation is $\sigma = \pi/\sqrt{6b}$. As n increases, the distribution remains identical in form, but moves to the right of a distance proportional to $\ln(n)$. It may be argued that we do not know the activity at a facility that attracts trips. For our purpose, it is in fact only necessary to assume that the proportion of trips ending up in facility j , $T_j'n$, is proportional to W_j , activity at facility j : $T_j'n = c'W_j$ where c' is a proportionality constant. Hence for trips to facility j , $\Phi(v) = \exp[-c'W_j e^{-b(v - \bar{v})}]$.

We can now calculate the surplus (or net benefit) offered to a trip maker from subarea i by the optimal trip to facility j . We define this surplus as the difference between the probabilistic utility v (the gross benefit of making the trip) and a deterministic trip cost C_{ij} incurred in making the trip. C_{ij} is a generalized cost incorporating direct payments, time costs, and so forth. The surplus is therefore given by $S_{ij}' = v_i - C_{ij}'$, and by substitution, we can obtain the probability that the surplus will attain any particular value S' :

Figure 2.33 PROBABILITY-DENSITY FUNCTION FOR THE UTILITY OF THE BEST TRIP



SOURCE: Cochrane (1975). Reprinted with permission.

$$\Phi_{ij}(S') = \exp[-c'W_j e^{-b(s' - \bar{v} + C_{ij})}] \quad (2.42)$$

where $\Phi_{ij}(S')$ is the cumulative distribution function of the surplus accruing from the preferred (optimal) trip between subarea i and facility j . Our basic assumption throughout is that a trip maker will choose the trip from his origin subarea that maximizes personal surplus. The probability that this trip from subarea i will be to facility j is the probability that the highest surplus offered by a trip possibility in facility j is greater than the highest surplus offered by any other facility. This probability is given by

$$\int_{-\infty}^{\infty} \phi'_{ij}(S') \left[\prod_{r \neq j}^J \Phi_{ir}(S') \right] dS' \quad (2.43)$$

This equation considers all the joint probabilities that “the surplus resulting from the trip to facility j has a value in the neighborhood of S' ($\Phi_{ij}(S')$) and that “the surplus resulting from a trip to another facility is less than S' .” Integrating from $-\infty$ to ∞ assumes that the trip will always be made even if the surplus is negative. However, if the cost determines which trip is made rather than whether a trip is made at all, the probability of the surplus being negative is very low and we can approximate with these limits of integration, which is simpler computationally.

Equation 2.43 can be rewritten as

$$\int_{-\infty}^{\infty} \left[\frac{\phi'_{ij}(S')}{\Phi_{ij}(S')} \left(\prod_j \Phi_{ij}(S') \right) \right] dS' \quad (2.44)$$

Differentiating Equation 2.42, we obtain

$$\phi'_{ij}(S') = bc'W_j \exp[-b(S' - \bar{v} + C_{ij}) - bW_j e^{-b(-\bar{v} + C_{ij})}] \quad (2.45)$$

Substituting Equations 2.42 and 2.45 into Equation 2.44, we obtain

$$\frac{W_j e^{-bC_{ij}}}{\sum_j W_j e^{-bC_{ij}}}$$

which is the same as the gravity model of Huff, as indicated in Equation 2.3, except a power function of travel time is now replaced by a negative exponential function of generalized spatial cost. Since the total number of trips originating from subarea i is V_i , the expected number of trips V_{ij} from subarea i to facility j is

$$\frac{V_i W_j e^{-bC_{ij}}}{\sum_j W_j e^{-bC_{ij}}} \quad (2.46)$$

which is the customary form of the singly constrained gravity model. We can calculate the total surplus arising from the trips actually made. The calculation uses methods unfamiliar outside statistics (see Cochrane [1975] for derivation):

$$\frac{1}{b} \sum_i V_i \left[0.577 + \ln \left(c' e^{b\bar{v}} \sum_j W_j e^{-bC_{ij}} \right) \right] \quad (2.47)$$

We are normally only interested in the change in surplus resulting from a change in trip costs from C^0 to C' , which can be represented by

$$\frac{1}{b} \sum_i V_i \ln \left[\frac{\sum_j W_j e^{-bC'_{ij}}}{\sum_j W_j e^{-bC^0_{ij}}} \right] \quad (2.48)$$

as shown in Equation 2.17 and illustrated in Figure 2.9.

Example

With the appropriate trip-utility, i.e., $W_j = 1$, and a single trip origin $V_i = 1$, Equation 2.46 can be simplified to read $\theta_{ij} = \exp(-bC_{ij}) / \sum_j \exp(-bC_{ij})$, Equation 2.47 becomes $S'_i = \frac{1}{b} \ln \sum_j e^{-bC_{ij}}$ and Equation 2.48 becomes

$$\frac{1}{b} \ln \left[\frac{\sum_j \exp(-bC'_{ij})}{\sum_j \exp(-bC^0_{ij})} \right]$$

Notice that if the travel-choice set has only one option, the summation sign vanishes and $S'_i = C_i = \bar{v}$. Suppose $b = 0.2$, $C_{i1} = 5$ and $C_{i2} = 8$ for the base-year and $C_{i1} = 5$, $C_{i2} = 8$ and $C_{i3} = 12$ for the forecast year after an accessibility improvement. These three expressions can be evaluated as shown in Table 2.4 and Table 2.5. The second expression S'_i —representing the utility or benefit (actually a disutility

Table 2.4 SAMPLE BENEFIT MEASURES BEFORE ACCESSIBILITY IMPROVEMENT

C_{ij}	b	$\exp(bC_{ij})$	θ_{ij}	\bar{v}	S_i'
$C_{i1} = 5$	0.2	0.3679	0.6457	6.0629	2.8126
$C_{i2} = 8$		0.2019	0.3543		
Total		0.5698	1.0000		
$C_{i1} = 5$	0.6	0.0498	0.8581	5.4257	4.7450
$C_{i2} = 8$		0.0082	0.1419		
Total		0.0580	1.0000		

SOURCE: de la Barra (1989). Reprinted with permission.

or dis-benefit in this case) from origin i —is evaluated at 2.8126 for the base-year, and 2.0738 for the forecast-year. The third expression, representing the difference in benefit attributable to accessibility improvement, is evaluated at $(1/0.2) \ln [0.6605/0.5698] = 0.7386$ (de la Barra 1989).

If $b = 0.6$, the surplus from origin i is evaluated at $S_i' = 4.7450$ for the base-year and 4.7227 for the forecast-year. The consumers-surplus increase is now 0.0228 (instead of 0.7386.) Remember that the $b = 0.2$ represents a low-sensitivity group while $b = 0.6$ a high-sensitivity group, where sensitivity in this case refers to responsiveness to cost. Thus the lower sensitivity group perceives a lower disutility from the same travel choice set when compared with the high-sensitivity group (2.81 against 4.75 in the base-year). For the consumer-surplus increase, the low-sensitivity group clearly benefits more from the accessibility improvement

Table 2.5 SAMPLE BENEFIT MEASURES AFTER ACCESSIBILITY IMPROVEMENT

C_{ij}	b	$\exp(bC_{ij})$	θ_{ij}	\bar{v}	S_i'
$C_{i1} = 5$	0.2	0.3679	0.5570	6.8772	2.0738
$C_{i2} = 8$		0.2019	0.3057		
$C_{i3} = 12$		0.0907	0.1373		
Total		0.6605	1.0000		
$C_{i1} = 5$	0.6	0.0498	0.8472	5.5092	4.7227
$C_{i2} = 8$		0.0082	0.1401		
$C_{i3} = 12$		0.0008	0.0127		
Total		0.0588	1.0000		

SOURCE: de la Barra (1989). Reprinted with permission.

(0.7386 versus 0.0228). These results show the importance of these surplus indicators in evaluating policy options. Traditionally, transport-related projects have been evaluated with a cost and time criterion, assuming that the preferred project will be the one producing the least of the average-travel-cost \bar{v} , where $\bar{v} = \sum_{ij} \theta_{ij} C_{ij}$. The numerical example above shows that this is clearly a fallacy— \bar{v} has increased from 6.06 to 6.88 and from 5.43 to 5.51, respectively, after accessibility improvement!

Using consumers' surplus, accessibility improvement will always produce benefits, however small, and these benefits will not be the same throughout various population groups. It can be seen, for example, that for the population with a low sensitivity to cost, the percentage of trips destined for the nearest zone, corresponding to $C_{ij} = 5$, is 0.6457. By contrast, for the population with a high-sensitivity to cost, the percentage rises to 0.8581. As a result, the average-cost \bar{v} paid by the high-sensitivity group will be lower than that of the low-sensitivity group (5.43 against 6.06). In the forecast-year (after accessibility improvement), 14 percent of the low-sensitivity group can now access the distant zone 3, against only 1 percent of the high-sensitivity group. Correspondingly, the average-cost \bar{v} of the former group rises from 6.06 to 6.88, while the latter group only moves from 5.42 to 5.51. The average utility indicators S'_i show in both cases an improvement when the new accessibility option is introduced, but they also show that the low-sensitivity group benefits more, because the dis-utility moves from 2.81 to 2.07 while the high-sensitivity group hardly moves from 4.75 to 4.72. Hopefully, this numerical example drives home the usefulness of interpreting the gravity model in terms of economic benefits.

B. The Doubly Constrained Model

Aside from a fixed number of trips originating from i , the doubly constrained gravity model also restricts the number of trips ending in j . This model is appropriate for work trips where the number of trips emanating from the origin residential subarea every morning is perfectly inelastic, and these trips are heading toward employment centers that have a specific number of jobs, at least in the short run. If there is no constraint on trip ends, there will be some employment centers j in which the number of unconstrained trip ends will exceed the number of jobs available. We assume that under these conditions competition will lead to the jobs being taken up by those trips for which the surplus available is greatest. This will occur either because the utilities of the set of trip ends are bid down or because the costs are bid up. In either case we may represent the effect as the addition of an extra cost r_j to the trip, these additional costs are set such as to restrict demand to the jobs available.⁸

We then rewrite Equation 2.42 as

$$\Phi_{ij}(S') = \exp[-c' W_j e^{-b(S' - \bar{v} + C_{ij} + r_j)}].$$

Substituting in Equation 2.44 and integrating as before, we obtain the probability of a trip ending up in employment center j :

$$\frac{W_j e^{-b(r_j + C_{ij})}}{\sum_j W_j e^{-b(r_j + C_{ij})}} = \frac{W_j e^{-br_j} e^{-bC_{ij}}}{\sum_j W_j e^{-br_j} e^{-bC_{ij}}} \quad (2.49)$$

The number of trips from i to j is correspondingly

$$V_{ij} = V_i \frac{W_j e^{-br_j} e^{-bC_{ij}}}{\sum_j W_j e^{-br_j} e^{-bC_{ij}}}$$

where r_j is the calibration constant chosen such that $\sum_i V_{ij} = V_j = c' W_j$ for all j as mentioned. It is clear that this model is equivalent to the conventional doubly constrained model

$$V_{ij} = V_i \frac{W_j a_{j0} e^{-bC_{ij}}}{\sum_j W_j a_{j0} e^{-bC_{ij}}}$$

where $a_{j0} = e^{-br_j}$ with both a and r representing a calibration constant. A numerical example of the doubly constrained gravity model is found in Chapter 3. The change in surplus resulting from a change in trip costs is given by

$$\frac{1}{b} \sum_i V_i \ln \left[\frac{\sum_j W_j e^{-br_j} e^{-bC'_{ij}}}{\sum_j W_j e^{-br_j} e^{-bC_{ij}}} \right]. \quad (2.50)$$

In order to balance the number of trip destinations with the number of origins over the entire area, some of the additional facility costs r_j will be positive and some will be negative. These values will result in a_{j0} 's less than and greater than one respectively. It should also be noted that the surplus expression represents the benefit received solely by trip makers.

C. The Unconstrained Model

The unconstrained model is the most difficult of the gravity models discussed so far, where the trip generation at origin is modeled in addition to trip distribution. A partially constrained model is suggested by Cochrane (1975) in which it is assumed that there exists an upper limit to the number of trips generated by any subarea—as the trip costs rise, some of the trips are no longer made. When integrating Equation 2.43 above, we took the limits of integration from $-\infty$ to ∞ . The low value was used because when the distribution of maximal surplus is very much greater than zero the probability of a negative value of surplus is negligible and we can obtain a simple integral by using these limits. This assumption implies that the primary economic force bringing about trip making is stronger than those that decide the choice between destinations. If this is not the case, we should integrate more precisely between limits of 0 and ∞ . This implies that the trip maker decides not to make even the optimal trip if the surplus is not positive. Where the utility of the trip is only of the same order as the cost, this is an important consideration. Certain social and recreational trips are likely to come into this category, although trips such as work trips do not. More will be said about this in the “Location-Allocation” chapter of Chan (2005).

Integrating Equation 2.43 between the new limits leads to

$$[1 - \exp(-b' \sum_j W_j e^{-bC_{ij}})] \frac{W_j e^{-bC_{ij}}}{\sum_j W_j e^{-bC_{ij}}} \quad (2.51)$$

where $b' = -c'e^{b\bar{v}}$. Trips executed V_{ij} can be expressed in terms of this unconstrained model by

$$V_{ij} = V_i(W_i, W_j, b', b, C_{ij})\Theta(W_j, b, C_{ij}) \quad (2.52)$$

where V_i is the trip-generation term and Θ is the trip distribution term. Each of these two terms can be equivalenced to Equation 2.51 by setting

$$V_i = W_i[1 - \exp(-b'\sum_j W_j e^{-bC_{ij}})]$$

and

$$\Theta_{ij} = \frac{W_j e^{-bC_{ij}}}{\sum_j W_j e^{-bC_{ij}}}$$

The trip-generation term constrains the total trips made in response to increases in the cost of trip making. Hence, if costs rise on particular links, the total number of trips changes in accordance with the trip-generation term to a certain limit, and the allocation of trips among destinations changes in accordance with the gravity trip-distribution term meanwhile. Again, we will further develop this model in the "Location-Allocation" chapter of Chan (2005).

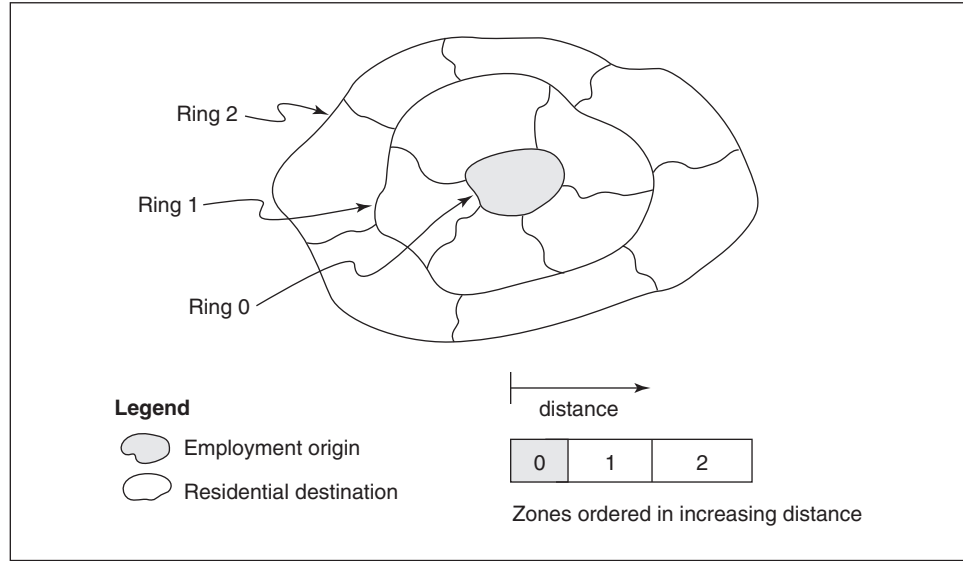
D. The Intervening Opportunity Model

Besides the gravity model, another common spatial allocation model is the intervening opportunity model (IOM). The IOM is based on a probabilistic formulation, which states that the probability, dP , that a trip will terminate in a destination is the joint probability that no termination point has been found among the total number of opportunities n visited so far and that the trip ends up in the current destination which offers an additional dn number of opportunities: $dP = [1 - P(n)] L' dn$. Here $P(n)$ is the probability that a termination point is found in the volume of destinations n , and L' is a constant probability that the subarea visited is in fact the termination point for the trip. Solving the differential equation for $P(n)$, the probability of finding a termination point in the n subareas visited is $P(n) = 1 - e^{-L'n}$. The expected number of trips from i , V_i , that will terminate in j , V_{ij} , is obtained by multiplying the total number of trips originating at i by the probability that the trip will terminate amid the n_j additional opportunities found in subarea j : $V_{ij} = V_i [P(n + n_j) - P(n)]$. Substituting the value of $P(n)$ in the above equation, the usual form of the IOM is

$$V_{ij} = V_i [e^{-L'n} - e^{-L'(n + n_j)}] \quad (2.53)$$

The basic theory of IOM states that (a) all opportunities are ordered by increasing distance from the origin and (b) the probability of an activity to be located at a particular destination is equivalent to a series of Bernoulli trials, where an activity is more likely to be located closer by than further away, everything else being equal. Thus in the residential location example in Figure 2.34,

Figure 2.34 DEFINITION OF OPPORTUNITIES IN THE INTERVENING OPPORTUNITY MODEL



- the probability of locating in destination 0 = L'
- the probability of locating in destination 1 but not in destination 0 = $L'(1-L')$; and
- the probability of neither locating in destinations 0 nor 1 but locating in destination 2 = $L'(1-L')(1-L')$.

In this example, there are five residential zones at a certain distance away from the employment zone, and there are seven zones yet further away. Here the number of zones within annular ring 0, 1, and 2 are $n_0 = 1$, $n_1 = 5$ and $n_2 = 7$. The zones are identified only by the annular ring in which they are located and all zones are assumed to be of equal size to denote that each offers the same residential opportunities. Alternatively, one can think of the destinations being ordered in increasing distance from the employment origin, each with 1, 5, and 7 opportunities respectively as shown in the lower part of the figure. If the probability of residential location in a zone, L' , is $1/2$, we can compute the relative frequency of residential activity distribution as

- percentage of population living in origin 0 = $e^0 - e^{-(1/2)(1)} = 0.390$
- percentage of population living in destination 1 = $e^{-(1/2)(1)} - e^{-(1/2)(6)} = 0.556$
- percentage of population living in destination 2 = $e^{-(1/2)(6)} - e^{-(1/2)(13)} = 0.048$

and so on.

The simple numerical example illustrates not only the computational mechanics of Equation 2.53, but also the problem of calibration. For example, we observe that assigning the value of $1/2$ to L' is merely arbitrary; its value needs to be calibrated from available trip-length-frequency data. Second, defining residential opportunity as the physical land area may be convenient, but a more workable

definition is likely to be problem specific and requires more effort. Finally, it is noted that the population allocation percentages up to the second annular rings do not add up to 100 percent. But if one considers additional annular rings ad infinitum, the sum of the percentages has to be unity according to Equation 2.53. Some practitioners prefer this model on the grounds that it can be developed from a defined set of statistical assumptions. Others have been concerned by the fact that the IOM has no intrinsic cost elements, and in particular does not distinguish the case where the subsequent opportunity is marginally more distant.

Curiously, it is possible to derive the IOM as a special case of the gravity model. We derive these models by assuming a relationship between the cost of transport between two points and the number of intervening opportunities. If we assume this to be of a power form:

$$n = b''[C_{ij}]^\beta \quad (2.54)$$

then $C_{ij} = [b'']^{-1/\beta} n^{1/\beta}$ where travel cost is not a function of distance as alluded to previously. In the singly constrained gravity model, we can write

$$V_{ij} = V_i \frac{n_j \exp(-b[b'']^{-1/\beta} n^{1/\beta})}{\sum_j n_j \exp(-b[b'']^{-1/\beta} n^{1/\beta})} \quad (2.55)$$

Substituting $b[b'']^{1/\beta} = b_0$

$$V_{ij} = V_i \frac{n_j \exp(-b_0 n^{1/\beta})}{\sum_j n_j \exp(-b_0 n^{1/\beta})}$$

and using the incomplete gamma function $\Gamma[x, y]$, Cochrane (1975) evaluated Equation 2.55 as

$$V_{ij} = V_i \left\{ \frac{\Gamma'[\beta, b_0(n + n_j)1/\beta] - \Gamma'[\beta, b_0 n^{1/\beta}]}{\Gamma'[\beta, b_0 n'^{1/\beta}]} \right\} \quad (2.56)$$

where n' is the total number of opportunities in the study area. The gamma function can be considered as a set of related functions of the second variable, the particular function to be used being indicated by the first variable, which in this case is β . If the number of opportunities is directly proportional to the cost as indicated in Equation 2.54, β is equal to one and the incomplete gamma function becomes the negative exponential function. We then obtain

$$V_{ij} = V_i \left[\frac{e^{-b_0 n} - e^{-b_0(n + n_j)}}{1 - e^{-b_0 n'}} \right]$$

If n' is large, the usual form of IOM results: $V_{ij} = V_i [e^{-b_0 n} - e^{-b_0(n + n_j)}]$. This derivation illustrates a very important concept in the analysis of spatial-temporal information. Through spatial cost transformation, apparently unrelated models can be equivalenced. We will have many other examples later on in this book and in Chan (2005) to illustrate this point.

IX. CONCLUDING REMARKS

In this chapter we have reviewed many of the basic economic concepts of facility location and activity allocation. We saw that the determination of spatial patterns—both in discrete facility locations and continuous land-use developments—can be explained in a set of common terms. These common constructs range from median to center models, from input-output analysis to the gravity model—all developed from basic economic concepts such as utility theory. Modern-day econometrics also allows empirically based approaches to be used to forecast future activity patterns. This is performed independent of the classic economic concepts, as illustrated in the interregional demographic projection section. In a fairly readable manner, it illustrates the basic building blocks of spatial-temporal information. To be sure, analysis of spatial-temporal information involves not only economic or econometric techniques, but the well-established economic concepts are convenient and familiar points of departure for many who work in this field.

In the next few chapters, we will provide the ways and means to further operationalize some of these concepts. In Chapter 3, we lay out the statistical procedures; while in Chapter 4, we outline the optimization algorithms. These techniques help to implement what were up to now theoretical constructs in terms of solid operational procedures. Recent advances in both descriptive and prescriptive tools allow us to realize some of the goals that our predecessors can only dream of. We then introduce a more recent paradigm for location decisions, multi-criteria decision making, which departs from traditional economics in several ways. First, it is behaviorally based rather than structurally based, complete with its own version of multi-attribute utility theory. Second, it broadens our concepts of ranking locations and shows that some counterintuitive results regarding transitivity and intransitivity among candidate sites may occur. For example, we demonstrate that site *A* preferred to site *B*, and site *B* preferred to site *C* does not necessarily mean site *A* is preferred to site *C*. Such recent advances in behavioral and mathematical sciences allow for a more innovative approach to modeling spatial decisions in general. It is one of our objectives to report these exciting developments here in this volume.

X. EXERCISES

Self-Instructional Module: PROBABILITY

(to be found on the attached CD/DVD)⁹

An understanding of probability is important to the decision maker. Many decisions must be based on predictions of future events. Inevitably, the prediction of future events has uncertainties and probable errors. An example is population projection, as discussed in Chapter 2 of this text. An understanding of probability concepts helps the decision maker to appreciate the significance of such uncertainties and probable errors.

This activity module is divided into three sections. The *first* section covers some of the theories of probability. The *second* section covers some rules of counting. Finally, the *third* section builds upon the first and second sections and illustrates with some interesting examples.

By the end of this exercise, the student

- (a) would be familiar with these concepts: sample space, events, union and intersection of events, empirical or frequency probability, subjective probability, and permutation.
- (b) would have seen some useful application of these concepts.

This module serves to introduce or review the fundamental probability concepts, which allows an understanding of what is *information* and *imperfect information*. An example of imperfect information is given in Chapter 2, in connection with locational competitions. Naturally, probability is required for the discussions in Chapter 3, where a number of descriptive analysis tools, including simulation, subjective probability, curve fitting, and information theory are formally discussed.

Probability is a prerequisite for an understanding of statistics, a basic building block of analytics. As such, it serves as an excellent introduction to a subsequent self-instructional module on Probability Distribution and Queuing. It is also a prerequisite for the Appendices entitled "Review of Statistical Tools" and "Review of Markovian Processes".

Problem 1: Gravity Model

The most common theory to explain spatial interaction is the Gravity Model, which states in mathematical terms the relationships between "activity at zone j " and "activity at origin zone i ," as governed by the "spatial cost between them:"

- Employment at zone $i = E_i$
- Spatial cost between i and $j = d_{ij}$
- Proportion of activities from origin i that end up in destination $j = \Theta_{ij}$

We can then write

$$E_i \Theta_{ij}(d_{ij}) = E_i \frac{N_j/d_{ij}^2}{\sum_k N_k/d_{ik}^2}$$

Correspondingly, populations N_j are to be distributed among the urban area according to

$$N_j = \sum_i E_i \Theta_{ij}(d_{ij})$$

which, when written for zone 2 in a city of three zones assumes the form

$$N_2 = E_1 \Theta_{12}(d_{12}) + E_2 \Theta_{22}(d_{22}) + E_3 \Theta_{32}(d_{32})$$

Here,

$$\Theta_{12}(d_{12}) = \frac{N_2/d_{12}^2}{N_1/d_{11}^2 + N_2/d_{12}^2 + N_3/d_{13}^2}$$

and so on.

For the following study area:

From/To	Zone 1	Zone 2	Zone 3
Zone 1	2	8	6
Zone 2	8	3	4
Zone 3	6	4	3
Base-yr pop: N_j	480	870	1020

$$\Theta_{12}(d_{12}) = \frac{870/8^2}{480/2^2 + 870/8^2 + 1020/6^2} = 8.3950 \times 10^{-2}$$

- (a) Please calculate $\Theta_{11}(d_{11})$ and $\Theta_{13}(d_{13})$ and verify that the sum of $\Theta_{11}(d_{11})$, $\Theta_{12}(d_{12})$ and $\Theta_{13}(d_{13})$ adds up to the numerical value of unity.

Let us say the total employment in zone 1 is projected to be 500. Accordingly, it means that $E_1\Theta_{12}(d_{12}) = 500 \times 8.3950 \times 10^{-2} = 41.975$, or 42 workers from zone 1 will live in zone 2. In order to forecast the *total* number of workers living in zone 2, however, two additional number are needed. They are $E_2\Theta_{22}(d_{22})$ and $E_3\Theta_{32}(d_{32})$. The sum of the three numbers is the total number of employees living in zone 2, in accordance with the equation $N_j = \sum_i E_i \Theta_{ij}(d_{ij})$.

- (b) Please calculate the number of employees living in zones 1 and 3.

Similarly, retail employment is located vis-a-vis people's residential choice. The probability that a shopping center will be located at zone 2, given a residential location at zone 1, is given by this formula

$$\frac{E_2^R/d_{12}^2}{E_1^R/d_{11}^2 + E_2^R/d_{12}^2 + E_3^R/d_{13}^2}$$

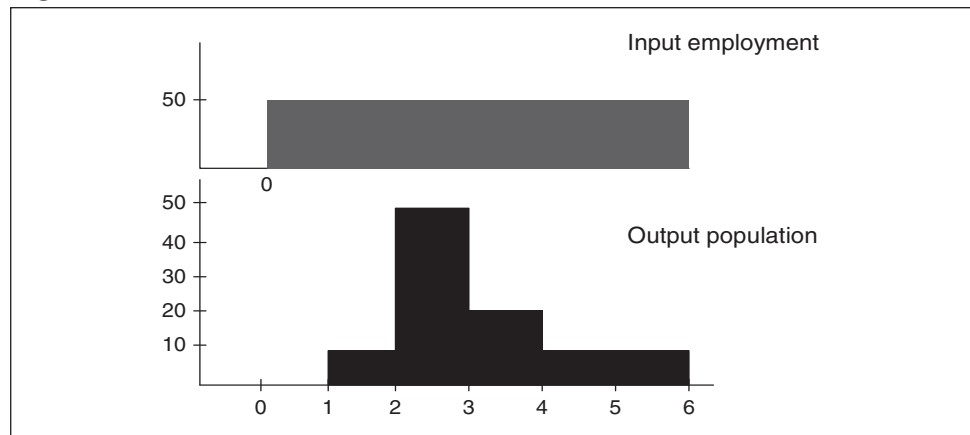
Problem 2: Further Discussions on Forecasting

It is commonly observed that population migration follows employment, albeit with a time lag. Here, we wish to estimate the distribution of population over the next six time periods following the introduction of employment (Chan 2005). The following shows the time lag for the dependent population to move into town:

- ☐ 0% of the population moves in during period 1 when employment is made available,
- ☐ 10% of the population moves in during period 2,
- ☐ 50% in period 3,
- ☐ 20% in 4,
- ☐ 10 in 5, and
- ☐ the remaining 10 in 6.

Such a time-lag relationship is also shown graphically below, where the number of jobs and the population size are expressed in ten's. For example, 50 jobs as shown in the top graph means actually 500 jobs, and a population of 10 actually means 100 in the lower graph (Figure 2.35):

Figure 2.35 INPUT–OUTPUT RELATIONSHIP



While 500 jobs are introduced in time period 1, subsequent jobs are available in varying quantities—300 in period 2, 900 in period 3 and so on. The same time-lag distribution is followed for subsequent employment introductions, as shown in the following Table 2.6. From the Table 2.6, it is clear that the employ-

Table 2.6 GROWTH IN ECONOMIC ACTIVITIES

Employment and population over time																						
period	emp*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	pop*
1	5																					0
2	3	5																				5
3	9	25	3																			28
4	6	10	15	9																		34
5	5	5	6	45	6																	62
6	7	5	3	18	30	5																61
7	3		3	9	12	25	7															56
8	4			9	6	10	35	3														63
9	8				6	5	14	15	4													44
10	7					5	7	6	20	8												46
11	8						7	3	8	40	7											65
12	1							3	4	16	35	8										66
13	3								4	8	14	40	1									67
14	3									8	7	16	5	3								39
15	7										7	8	2	15	3							35
16	3											8	1	6	15	7						37
17	6													1	3	6	35	3				48
18	4														3	3	14	15	6			41
19	1															3	7	6	30	4		50
20	3																7	3	12	20	1	43
emp		50	30	90	60	50	70	30	40	80	70	80	10	30	30	70	-	-	-	-	-	
*Figures are shown in units of 10's. For example, 50 jobs (instead of 5 jobs) are introduced to the study area in period 1.																						

ment can be gleaned from the bottom row. Each introduction of employment triggers in-migration of population, following the given time-lag distribution. The row sums amount to the total population in the study area for each time period, which are summarized in the right-most column in the Table 2.6. The following Table simply shows the employment and population series side-by-side, as extracted from the master Table 2.6 above:

period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
emp	50	30	90	60	50	70	30	40	80	70	80	10	30	30	70	-	-	-	-	-
pop	0	5	28	34	62	61	56	63	44	46	65	66	67	39	35	37	48	41	50	43

Repeat the calculations for the same employment series, except that the time-lag distribution is now changed to

- ☐ 0% of the population moves in during period 1,
- ☐ 20% of the population moves in during period 2,
- ☐ 40% in period 3,
- ☐ 25% in 4,
- ☐ 15 in 5, and
- ☐ the remaining 10 in 6.

ENDNOTES

- ¹ In Section VIII of Chapter 4, we will discuss how to measure efficiency using Data Development Analysis, which is based on non-dominated solutions to more than one cost criterion.
- ² A full explanation of Hoyt's theory is given in Section II.B of Chapter 6.
- ³ Figure 2.10 shows the tradeoff between the quantity of transportation and housing consumed in an indifference curve. On an indifference curve, a family sacrifices travel for better housing or vice versa for a given income. The tangency of the income/budget straight line and the indifference curve is the consumption level of the family.
- ⁴ The series expansion for $\ln(1 + x)$, where $-1 \leq x \leq 1$, is $x + x^2/2 + x^3/3 + \dots$
- ⁵ Much of the discussion in this section is taken from Vickerman (1980).
- ⁶ These two cases will be analyzed in detail in later chapters when we construct models of market equilibrium.
- ⁷ Recursive programming is explained in Appendix 3. Chan (2005) also illustrated application of recursive programming in his "Location-Routing Models" chapter. A software example is included on the attached CD/DVD under the RISE folder.
- ⁸ Chan (2005) discussed alternate ways to effect this reallocation in his "Lowry-based Models" and "Bifurcation and Disaggregation" chapters. The readers may also wish to experiment with the software on the attached CD/DVD under the LOWRY and YI-CHAN folders.
- ⁹ The answer to this module is attached at the end of this text book.

REFERENCES

- Ahivuv, N.; Berman, O. (1988) *Operations manual of distributed service networks—A practical quantitative approach*. New York: Plenum Press.
- Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Cambridge, Massachusetts: Harvard University Press.
- Alonso, W. (1970). "Equilibrium of the household." In *Urban analysis: Readings in housing and urban development*, edited by A. N. Page and W. R. Siegfried. Glenville, Illinois.: Scott, Foresman and Co., 168–177.
- Alonso, W. (1960). "A theory of the urban land market." *Paper and Proceedings of the Regional Science Association* 6:149–157.
- Anderson, S. P.; Never, D. J. (1991). "Cournot competition yields spatial agglomeration." *International Economic Review* 32:793–808.
- Bonsall, P. W.; Shires, J. D. (2006). "Employers' expectations for commuting & business-related travel in an ICT-Rich environment." *Transportation Research Record*, 1977:268–276.
- Chapin, F. S.; Kaiser, E. J. (1979). *Urban land use planning*, 3rd ed. Urbana, Illinois.: University of Illinois Press.
- Chan, Y. (2005). *Location, transport and land-use: Modelling spatial-temporal information*. Berlin and New York: Springer.
- Cochrane, R. A. (1975). "A possible economic basis for the gravity model." *Journal of Transport Economics and Policy* 9:34–49.
- Dahlman, C. J. (1988). "The problem of externality." In *The theory of market failure*, edited by T. Cowen. Fairfax, Virginia: George Mason University Press, 168–177.
- de la Barra, T. (1989). *Integrated land use and transport modelling: Decision chains and hierarchies*. Cambridge, England: Cambridge University Press.
- Dickey, J. W. (1983). *Metropolitan transportation planning*, 2nd ed. New York: McGraw-Hill.
- Dorau, H. B.; Hinman, A. G. (1928). *Urban land economics*. New York: The Macmillan Co.

- Friedrich, C. J., ed. (1929). *Alfred Weber's theory of the location of industries*. Chicago: Chicago University Press.
- Hakimi, S. L. (1990). "Location with spatial interactions: Competitive locations and games." In *Discrete location theory*, edited by P.B. Mirchandani and R. L. Francis. New York: Wiley-Interscience, 439–478.
- Hakimi, S. L. (1964). "Optimal location of switching centers and the absolute centers and medians of a graph." *Operations Research* 12:450–459.
- Huff, D. L. (1962). Determination of intraurban retail trade areas. Real Estate Research Program. University of California at Los Angeles. Los Angeles.
- Jha, K. Demographic models. Working Paper. Department of Civil Engineering. Pennsylvania State University. University Park, Pennsylvania.
- Krueckeberg, D. A.; Silver, A. L. (1974). *Urban planning analysis: Methods and models*. New York: Wiley-Interscience.
- Lakshmanan, T. R.; Hansen, W. G. (1965). "A retail market potential model." *Journal of the American Institute of Planners* 31, No. 2:134–143.
- Leontief, W. W. et al. (1953) *Studies in the structure of the American economy*. New York: Oxford University Press.
- Lund, J. R.; Mokhtarian, P.L. (1994). "Telecommuting and residential location: Theory and implications for commute travel in the monocentric metropolis." *Transportation Research Record* 1463, pp. 10–14.
- Mai, C-C. (1986). "Random input transport rate and optimum location of the firm." *Transportation Planning Journal* (Taiwan, Republic of China) 18:311–331.
- Marsh, M. T.; Schilling, D. A. (1994). "Equity measurement in facility location analysis: review and framework." *European Journal of Operations Research*, 74, No. 1:1–17.
- Mokhtarian, P. L.; Meenakshisundaram, R. (1999). "Beyond tele-substitution disaggregate longitudinal structural equation modeling of communication impacts." *Transportation Research* 7:33–52.
- Mokhtarian, P. L.; Varma, K. V. (1998). "The trade off between trips and distance traveled in analyzing the emissions impacts of center-based telecommuting." *Transportation Research* 3:419–428.
- Nelson, J. (1971). "An interregional recursive program model of production, investment, and technological change." *Journal of Regional Science* 11: 33–47.
- Newman, E. E. (1972) Economic concepts and models. Working Paper. Department of Civil Engineering. Pennsylvania State University. University Park, Pennsylvania.
- Papageorgiou, Y. Y. (1990). *The isolated city state: An economic geography of urban spatial structure*. New York: Routledge (Chapman and Hall).
- Van Lierop, W. (1986). *Spatial interaction modelling and residential choice analysis*. Hants, England: Gower.
- Vickerman, R. W. (1980). *Spatial economic behaviour: The microeconomic foundations of urban and transport economics*. New York: St. Martin's Press.
- Yeates, M.; Garner, B. (1980). *The North American city*, 3rd ed. San Francisco: Harper and Row.

Location Theory and Decision Analysis

Analytics of Spatial Information Technology

Chan, Y.

2011, XXXIII, 701 p. 280 illus. With CD-ROM., Hardcover

ISBN: 978-3-642-15662-5