A Consumption-Based Accessibility Index of Transportation and Land Use

by

Michael Scott Ramming

B.S., Civil Engineering
Washington University in St. Louis, 1991

Submitted to the Department of Civil and Environmental Engineering in
Partial Fulfillment of the Requirements for the Degree of

Master of Science in Transportation

at the

Massachusetts Institute of Technology

February 1994

Signature of Author_____
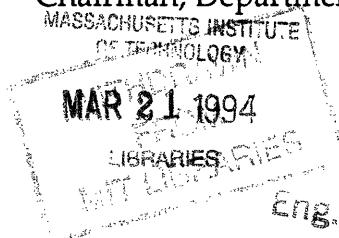Department of Civil and Environmental Engineering
5 November 1993

Certified by_____
David H. Bernstein
Assistant Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by_____
Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies

A Consumption-Based Accessibility Index of Transportation and Land Use

by

Michael Scott Ramming

Submitted to the Department of Civil and Environmental Engineering on
5 November 1993 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Transportation

Abstract

Accessibility is examined from the economic viewpoint of consumer surplus net of travel disutility. The theory for an accessibility index is derived from discrete choice theory and refined so that conventional data sources, such as the U.S. Decennial Census, may be used to estimate the parameters of such an index. An attempt is made to apply the accessibility methodology to data from the Boston metropolitan area to allow comparisons of possible transportation or land use-related government expenditures. The mode choice submodel reveals that the region's residents are particularly sensitive to out-of-vehicle travel disutility. Unfortunately, however, housing values alone are insufficient to provide statistically significant estimates of determinants of accessibility.

Thesis Supervisor: David H. Bernstein
            Title: Assistant Professor,
                      Department of Civil and Environmental Engineering

Acknowledgments

Many people have contributed to this thesis and thanks are due ...

David Bernstein for being interested in the relations between transportation and land use, agreeing to supervise this research effort, and having faith in my abilities to conduct this project in a timely, rigorous manner, even when I doubted those abilities;

Tom Piper for helping me see the grand scheme in the midst of my equations, and for his help in obtaining data from CTPS;

Val Southern, Arnie Soolman, Yong Chang, Mark Scannell, B. J. Mahal, and Steven Falbel at CTPS for providing travel data, and Kurt Schumann at the Central Massachusetts RPC for answering questions after normal business hours;

the sponsors of the Boston Globe Livable Region Conference, which funded this research, and then-Secretary of Transportation (last September, when this project was just forming) Richard Taylor for having the courage to support our attempt to improve the means by which people look at transportation systems;

Fred Salvucci and James "Kibo" Parry for continuing to remind me of the true reason behind transportation, and for sharing many great stories;

Ruth Bonsignore, Dan Turk, Lisa Cole, Anne Kinsella, Tilly Chang, Bill Cowart, Dinesh Gopinath, Milly Polydoropoulou, Oki Hatsgai, other members of the Livable Region Conference team, and classmates who have shared data and insights with me;

Bill Twomey of TAMS, Nagi, and Geoff Lauprête for their interest in my work and their desire to continue examining accessibility;

Moshe Ben-Akiva for discussions about the derivation and statistical properties of my model;

Jerry Rothenberg for discussions about the many complexities of modeling reality;

my friends and colleagues from the East-West Gateway Coordinating Council in St. Louis for giving me an introduction to the urban transportation planning process;

Charles River Associates for their support and patience in waiting for the results of this thesis, and Mark Hickman for the glowing recommendation which got me a job at CRA in the first place;

roommates Dongsu, Denis, Harrison, and Nick who put up with my unusual sleep schedule; sun, jpkirby, queef, tegeler, jjgraham, gagne, elabenur, eddietwo, danw, mdtanner, dalbader, and everyone else on Athena who put up with my silliness; Mom and Dad for supporting and believing in me, even when it would be weeks between calls home;

and of course caffeine in all its wonderful forms, including Necco Sky Bars, Constant Comment Tea, chocolate covered espresso beans, and Folgers Coffee Singles.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Summary

Many transportation projects list improving mobility as one of their goals: moving people faster, or moving more people, or both. Such criteria are easy to meet, as transportation professionals generally have an excellent understanding of vehicle capacities and flows. However, transportation planners and citizens alike have called for less attention on mobility — merely being able to move from place to place — and more attention on accessibility, which incorporates some notion of the benefits of what can be reached. Indeed, such an approach is justified any time, but is even more necessary during times of constricting resources.

1.1. The Need to Measure Accessibility

Accessibility is an important analysis and evaluation measure because it puts people's wants and desires as the paramount concern. An accessibility measure would help both politicians and transportation planners, since it looks at transportation as an investment which enables improvements in urban residents' quality of life. Accessibility can also be used to examine the effects of other "enabling" expenditures, such as development incentives, education reform, or crime-fighting programs.

This approach is in sharp contrast to traditional transportation planning measures, such as what road segments are over capacity or severely congested, or how many people ride a certain transit line Accessibility has the advantage of considering the transportation and land use systems together. In other words, a change to either system affects accessibility. Thus while traditional measures often lead planners to think exclusively about transportation improvements, accessibility should lead them to think about both. In short, instead of focusing on transportation characteristics such as level-of-service, accessibility focuses on the impacts on residents' life processes (that is, all activities important to urban dwellers, such as their jobs, their consumption of goods and services, and their ability to get sufficient health care and education).

We know that it is important not to consider travel solely in terms of mobility, or travel for travel's sake, since reactive "sizing" (e.g. widening a road just because it's overused now) completely neglects the effect that changes to the transportation network have on other systems such as land use. Transportation systems that try to serve demand retroactively are obsolete before they are built. Instead, we need to consider the decisions that motivate the travel: people need the ability to go places in order to buy the necessities of survival, and usually have income left over which they spend at other places to make life more enjoyable. The accessibility measure developed here has a behavioral and microeconomic framework that makes it more robust than traditional measures, which often use

arbitrary functions based on the sole consideration that those functions "fit" past travel best.

Perhaps the easiest way to understand the importance of an accessibility measure is with some examples. In a city with low accessibility, certain types of stores may have "geographic monopolies" since there are so few of these stores and the stores are hard to get to. But by improving accessibility (by improving transportation, or by reducing barriers to entry such as the ability to get large loans), the market will be made "contestable" — that is, it may appear at first glance to be a geographic monopoly, but stores aren't able to charge monopoly prices because either consumers would travel farther to other stores, or new stores would enter the market. Another benefit of geographic competition is that it will lead to better quality products being produced and a greater number of options available to citizens.

Another use of this accessibility measure is for examining how urban residents might like a certain activity to be distributed throughout the city. A policy-maker may be interested in providing low income housing. The analyst might determine that low income persons place a high priority on access to blue-collar and service jobs, supermarkets, fast-food restaurants, discount and hardware stores, quality public education, sports arenas, and places of worship, while placing lower priorities on access to sit-down restaurants, theaters, and department stores. Then the analyst could construct an accessibility index that reflects where low income people

would *like* to live. Information is readily available on where different socioeconomic groups do live and where vacant land or buildings are available. The analyst would "match up" the desire for low income housing with areas of opportunity, and therefore recommend where the creation of "enterprise zones" for development incentives might be most effective.

1.2. Problem Statement

In order to receive the full planning benefits of an accessibility measure, we need to carefully consider what aspects of life such a measure should reflect. An accessibility measure should deal explicitly with the fact that different people have different preferences. Different preferences are important for several reasons. Certainly the world would be a much more boring place if everyone was identical. But more importantly, differences are good because they allow people to get the most enjoyment out of what they like most, without getting in other people's way. Such "Jack Sprat" outcomes occur in many policy questions. The "rebuttable presumption" of the Boston Transportation Planning Review (BTPR) is one such situation: core-oriented growth is good for all since the core can easily be made accessible to both those who prefer the open spaces of the suburbs and those who enjoy the activity of the center city.

An accessibility measure should also address the microeconomic notion of scarcity. Because of constraints on time and money, travelers don't make

all the trips they could or even *would* like to make. Traditional measures are concerned only with trips actually made and ignore trips that people might like to make but haven't been able to. However, improvements to transportation affect not only existing trips, but also allow new trips to occur. Accessibility's "potential" travel definition helps us understand "latent demand" better, since accessibility reflects the *whole* travel demand curve instead of just one point on that curve. Considering potential travel is important because unchosen options *do* have value to people. Accessibility can look at the different types of preferences that urban residents have, and how the choices made by those people will change when their options change (from a certain type of shop opening or closing in such and such a location, for example).

Finally, an accessibility measure should be able to incorporate traditional comparisons of transportation projects with other important government expenditures and influences on travel behavior. Perhaps the most important motivation of travel is the spatial distribution of activities throughout a city or region. Equally important are taste differences among citizens who consume transportation services, housing, and countless other goods. The challenge, then, is to develop a model of accessibility which can take into account the complexities of human behavior, but yet is simple enough to be able to be estimated and applied using readily available data and computational equipment.

1.3. Outline of This Thesis

This study will begin by examining basic microeconomic and discrete choice theory and developing a specification flexible enough that planning biases need not be introduced for the sake of model simplicity. Then for specified purposes, the model can be reduced to allow more direct estimation and application, at the expense of not being able to examine as wide of a range of planning options.

Chapter 2 examines current literature on accessibility as it relates to performance of transportation networks and residence choice decisions. This chapter also reviews the econometric theory relevant to the approach of this thesis. Chapter 3 discusses some issues involved with developing a useful, operational measure of accessibility. Chapter 4 outlines the theoretical model, and chapter 5 describes the results of estimating such a model for the Boston region. Chapter 6 examines the implications of model results and offers some general policy suggestions.

Chapter 2

Review of Theoretical and Empirical Research

This research attempts to review the methods used by several fields in order to better understand how transportation and land use systems interrelate. Sociologists and behavioral scientists concentrate on the activities people would prefer to participate in; the distribution in time and space of those activities; and how that distribution, along with other perceptions of the activities, motivates people's actions. More traditional transportation planners tend to have a supply focus: they calculate travel times and costs, and measure land uses in terms of square footage and persons employed. Economists assume individuals act rationally, and therefore examine how these decision makers trade off attributes of various products to maximize their objectives. Each of these approaches is examined below. Next, the work of other researchers to combine each of these approaches into transportation and location choice models is examined. Finally, results of some empirical work suggesting the significance of accessibility is presented.

2.1. Sociological and Behavioral Framework

Urban passenger transportation is a result of people's behavior patterns. Thus, to understand accessibility, it is necessary to examine the spatial areas within which an urban dweller conducts various activities, the motivations

between the supply and demand for these activities, and the allocation of time among activities. There are several approaches to this problem:

Barrett (1974) develops a behavioral approach to residence choice based on the concepts of action space, awareness space, place utility, search behavior, and vacancy set:

> The 'action space' consists of the points and paths that the individual uses in his normal space patterns for a given period of time. ... In fact, the 'knowledge' of these actual places is filtered through personal, cultural, social and economic screens so that in reality the 'action space' is a perceived space. Therefore the term 'awareness space' appears to be a more appropriate term .... 'Place utility' is the degree of differentiation between the satisfaction of a person's present location and the perceived advantages of moving to another location (Barrett cites Wolpert, 1965). ... A fourth factor, 'search behavior' is the action taken on the part of the potential mover to acquaint himself with possible alternative locations. ... Since persons can only move into available places, the type of places and where they are located is spatially significant. Thus the term 'vacancy set' fluctuates (over time) in both size and location ....

The concepts of awareness space and place utility will be useful in developing a measure of accessibility in terms of the consumption possibilities associated with a location.

Chapin (1974) considers the aggregate supply and demand for activities and the hierarchical relationship of people to groups such as families, cliques, gangs, firms, governments, religions, and ethnicities. In Chapin's framework,

the demand for, or propensity to engage in, activities is dependent on "motivations and thoughtways predisposing action" and "roles and person characteristics preconditioning action." The supply of, or opportunity to engage in, activities is modeled as being dependent on the perceived availability of facilities or services, and the perceived quality of those facilities or services. Brand (1990) proposes a similar model where individual behavior is based on needs, resources, and information on opportunities. The resulting individual decisions are observed in aggregate travel and land use patterns.

Most of the research on this subject focuses on the amount of time that households allocate to general activities, but how this allocation process relates to transportation is still the subject of ongoing research. For example Hammer & Chapin (1972) examine the use of leisure time in Washington, D.C., but ignore the separation in locations of these activities. Ettema, Borgers, and Timmermans (1993) examine how constraints on activities affect the scheduling of activities (for example, a shopper may need to go to the bank in order to have sufficient cash for grocery shopping, and then return home quickly before frozen foods defrost), which in turn influences trip chaining and travel patterns within the day. Kunert (1993) argues that households have weekly cycles of trip-making behavior, and that one day time-allocation or travel surveys do not obtain sufficient information about trip generation rates.

We will examine how economic actors (individuals or households, and firms) allocate time and money resources to maximize consumption or profit, and how this allocation is affected (and in turn affects) the transportation network and land use patterns of a city. To illustrate, consider individuals attempting to maximize their satisfaction from consumption not only by choosing which products to buy, but where to buy these products. The choice of where to buy products depends both on the difficulties presented by the transportation network in reaching that place, and the opportunities available at the final destination. The next two sections explore how each component of this choice — components which we will call "travel disutility" and place utility — may be quantified in a meaningful manner.

Before examining measures of accessibility though, it useful to discuss what such a measure should attempt to incorporate. In some reports, the terms *accessibility* and *mobility* are used interchangeably to refer to the degree of separation of various human activities (ex. Morris, Dumble and Wigan, 1979). As an example of what is meant by *activities*, Hunt et. al. (1984) define accessibility in terms of the distance to and frequency of transit service, in order to examine the equity of service provision in northern New Jersey across income groups. Linneker and Spence (1992) approach accessibility from the viewpoint of manufacturers who rely on vans to deliver their products to market, and examine the changes in accessibility from the construction of a circumferential freeway around London.

In related literature, *mobility* is used to indicate a change in residence (ex. Butler et. al., 1969, and Simmons, 1968). Also, *accessibility* is sometimes used in the transportation field to refer to the lack of physical constraints on certain groups, particularly the disabled and elderly, which would otherwise make travel by a specified mode difficult or impossible. (An example of this use occurs in the Americans with Disabilities Act of 1990.)

Morris, Dumble and Wigan (1979) offer this distinction between mobility and accessibility:

> ... personal mobility is interpreted to mean the ability of individuals to move from place to place: this depends principally upon the availability of different moves of transportation, including walking.

> .... On one hand accessibility may be interpreted as a property of individuals and space which is independent of actual trip making and which measures the *potential* or *opportunity* to travel to selected activities. Alternately, it may be held that "proof of access" lies in the use of services and participation in activities, not simply in the presence of opportunities.

This study will adopt the convention that mobility refers to the physical and technological constraints on the choices of travel modes available. Accessibility will refer to the set of activities to which a person has the *potential* to travel, even if such a trip is not made. This assumes that people view an unused opportunity as better than (or at least as good as) no opportunity. However, discussion of how opportunities are valued will be deferred until a later section.

2.2. Measures of Accessibility and Travel Impedance

The simplest measure of accessibility is merely the distance or time separating two activities. However, even this simple concept can quickly become complex. For example, distances may be calculated as straight-line, network, or "block" distances. Barrett (1973) defines block distance, a proxy for network distance, as "the right angle distance of an equilateral triangle." Although the meaning of this definition is not completely clear, we can assume that block distance is straight-line distance times some factor, which would otherwise appear within the estimated parameter on straight-line distance. Different components of travel time (such as access time v. in-vehicle-time) are often weighted to reflect the relative comfort of each. Distance and time may be combined with other aspects of travel, such as cost or preference for a given mode, to develop a composite value called "generalized cost." If this composite value is expressed in time units rather than monetary units, it is generally called "impedance." The value may also be normalized for use in probabilistic discrete choice models, in which case it is generally called "disutility" and expressed in units called "utils" or simply "utility units."

Ingram (1971) refers to accessibility measures which reflect the separation between only two activities as "relative accessibility." The corresponding concept — the separation between one point and all others in a region — is called "integral accessibility." Relative accessibility is a useful concept only

when considering an activity which is highly centralized (such as a City Hall); or which is carried out in homogenous facilities (an example might be post offices which offer the same services), and therefore only the nearest one is relevant. However, virtually all privately-operated activities and many public sector activities take place in several locations and have varying levels of quality and types of products offered.

One way to use measures of relative accessibility to examine accessibility within a region is to construct "isochrones" which bound all the opportunities available within a certain travel time (or generalized cost) budget. The total number of opportunities within a given budget band can be expressed as

$$Accessibility_i = \sum_j A_j \delta_{ij}^N$$

where A_j = some measure of the opportunities at destination j (opportunity measures will be discussed later), $\delta_{ij} = 1$ if $t_{ij} \leq t_N^*$ and 0 otherwise, and $t_N^* = a$ travel time budget for isochrone band N (an example is multiples of 10 minutes of impedance).

Ideally, one would want to construct one isochrone which corresponds to the relevant travel budget for a given person. However, travel budgets likely vary among individuals and it is difficult to collect sufficient information, so typically, isochrones are constructed for a series of travel budget bands,

resulting in much more data complexity. Several numbers are necessary to express the integral accessibility of an area whereas we would like to use a single index. (For an illustration of the isochrone approach, see the examination of accessibility within Toronto by Dewees, 1978)

Lerman (1975) reduces the data complexity of the isochrone approach by using the *expected value* of travel time (disutility, etc.) as a measure of accessibility:

$$Acc_i = \sum_j t_{ij} P_{ij}\left(t_{ij}, A_j, Z_i\right)$$

where Acc_i = the expected value of the travel disutility of a trip made from origin i, Z_i = a vector of socioeconomic characteristics of trip maker at origin i, and $P_{ij}(\bullet)$ = the probability that a resident of origin i will make a trip to destination j. The function $P_{ij}(\bullet)$ is usually from a discrete choice model and can incorporate socioeconomic variables to reflect the relevance of opportunities to various trip makers. However, Lerman's model considers only travel time disutility, that is $P_{ij}(t_{ij}, A_j, Z_i) = P_{ij}(t_{ij})$. Also, one would prefer a measure of the opportunities, rather than the time to reach these opportunities, as a means of describing accessibility, so that accessibility would be directly proportional to the attractiveness of an area.

Another measure of accessibility, proposed by Hansen (1959), is based on declining attractiveness as activities become more distant:

$$Accessibility_i = \sum_j A_j f(t_{ij}),$$

where A_j = a measure of the opportunities at destination j, typically area, population or jobs, t_{ij} = travel time and cost between origin i and destination j, and $f(t_{ij})$ = a decreasing function in t_{ij}, such as $1/(\alpha+t_{ij}^{\beta})$ or $\exp(-\gamma\, t_{ij})$. One of the difficulties with this approach is that the parameters α, β, and γ have little theoretical meaning. Often an arbitrary form of $f(\bullet)$ is chosen so that when accessibility measures are incorporated into trip generation rates, a gravity trip distribution model emerges. (see Morris, Dumble and Wigan's review, 1979, of Niedercorn and Bechdolt, 1969) Also, this formulation make no adjustment for the socioeconomic relevance of the opportunities at each destination.

2.3. Hedonic Price Estimation of Product Attributes

Instead of the approaches above, we will examine how the attractiveness of a destination might be expressed in similar units as travel times and costs. Borrowing utility theory from economists provides a means to make these comparisons. Economists often view heterogeneous products as an inseparable package of varying quantities of homogenous "attributes." (Examples of attributes are blueness, sweetness, and absence of impurities.)

By recording customers' willingness to purchase related products, and measuring the attributes inherent in those products, the values that customers place on those attributes can be inferred. This procedure is known as hedonic price estimation. Ting (1971) gives guidelines for when an additive utility function is appropriate (that is, when preference for attributes are independent), and provides mathematical tools for manipulating choices among multiattributed products. Wallace and Sherret (1973) show how the value of qualitative attributes may be quantified by using surveys where respondents are asked to rank their satisfaction with the attributes of a product. Wallace and Sherret also argue that the demand and supply functions of a multiattributed product must be considered simultaneously in order to correctly identify the demand for individual attributes. Moorthy (1991) warns that the design of such a survey bias the estimated hedonic price of the product's attributes.

Within the urban transportation demand framework, we consider locations as a composite product of consumption goods which can be purchased there and of public goods, such as ambiance, noise, or crime, which are not consumed but remain as a characteristic of the area.

Butler (1977) uses hedonic prices to explain prices of rental and owner-occupied housing units within a metropolitan area and across 36 cities in the U.S. Unlike Wallace and Sherret, Butler argues that the hedonic price pertains only to demand and supply factors together. Therefore, including buyer (or seller) characteristics into the hedonic regression will introduce

simultaneous equation bias. Butler also examines the bias introduced by using only variables readily available from the census, rather than all applicable housing attributes, and concludes that the explanatory power of such a hedonic model is little affected by the omission of non-census variables. This result suggests cause for optimism that census data may be sufficient to examine the influence of accessibility on transportation and land use patters.

2.4. Residence and Commercial Location Models

Although accessibility has a large influence on people's behavior, it is not directly measurable in the same way that say the travel time between two points by a given mode can be measured. Instead, we must look at the outcomes of decisions known to be influenced by accessibility. For this project, we will examine how accessibility affects housing values as a means of estimating a useful accessibility measure. Economists such as Alonso (1964) and Wingo (1961) model interactions of the transportation and land use markets with other, more traditional markets. Alonso theorizes that different land uses (industries and residences of varying density) trade off purchases of accessibility to the CBD, land area, and a composite good. The land use which offers the greatest "bid rent" end up with control of a parcel of land, which explains why retail establishments and industries are located in the city centers, while residences prefer outlying areas. Wingo examines the labor market from the perspective that households trade off higher rents near the commercial center city with longer commutes from the suburbs.

"Economic rent theory" such as used by Alonso and Wingo often assume a city located on a flat, featureless plane which results in circular bands of different land use patterns. Experience shows that neither the assumption nor result is descriptive of real cities.

Two early models which relate transportation to land use are the EMPIRIC and DRAM models. The EMPIRIC model (see Brand, Barber and Jacobs, 1967) examines changes in sub regional shares of residential and industrial activities (households and jobs) in terms of the existing levels of those activities, the propensity and capacity of an area to attract certain types of development, the level of utilities service, and automobile and transit accessibility. Both models (this discussion follows one in Dickey, 1983) use a Hansen-type (exponential gravity) accessibility measure. In the DRAM model, attractiveness is assumed to be estimable from a multivariate regression on areas of current land uses and the distribution of residents' income within quartiles.

Lerman (1975) formulates residential choice as a joint selection from a discrete set of residence, auto ownership, and mode to work choices. Accessibility is expressed as the reciprocal of expected generalized travel cost to shopping for separate transit and auto modes. Location and travel-related decisions are modeled by a generalized extreme value formulation where utility is derived from accessibility and other variables.

Similarly, Weisbrod (1978) models residential choice as a joint decision with auto ownership, neighborhood, tenure type, and physical structure (ex. a semidetached unit) in a generalized extreme value formulation. Weisbrod also expresses accessibility in terms of expected generalized travel cost, but combines the measure over all modes by using the expected least generalized cost among all modes. (Ben-Akiva and Lerman (1985) show that when the stochastic disturbances of all modes' utilities are independent, the combined expected value reduces to the "logsum" of the modes' systematic utilities. This method will be discussed in more detail in chapter 4.)

Anas (1985) approaches the residential choice problem from an optimization standpoint by examining the simultaneous equilibria of transportation network congestion and the housing market. Since land owners attempt to minimize their generalized travel costs while maximizing their benefit from housing occupancy, accessibility is implicitly considered. Anas considers the case where only one mode, auto, is available, but his framework can be expanded to include multiple modes.

2.5. Empirical Results Suggesting the Importance of Accessibility

Researchers using models such as those described above and "Push-Pull" models of residence choice have found that accessibility is one of several important factors home buyers and renters consider. Butler et. al. (1969) examine opinions of householders who had contemplated changing or had recently changed residence to determine what factors (such as change in

family composition, change in income, or current housing falling into disrepair) influence people's decision to move. They conclude that the head of the household's place of employment is the most important activity that households seek to become closer to by moving. Accessibility to shopping centers, parks and playgrounds, and friends was also noted to be statistically significant in influencing the decision to move.

Barrett (1973) examines various aspects of the search and evaluation process that 380 households in the Toronto area encountered while finding a new house. Although Barrett notes that the *sample average* distance (when measured by straight-line and block distances) to both the primary and secondary work locations for the households decreased after moving, he is unable conclude if this occurred in general for *individual* households. A survey asked the recent movers what reasons influenced their selection of a residence (households were allowed more than one response). While the most common answer was "it was the best value for the money," a vague reply at best, accessibility was mentioned by several households. Fifty-two households (about 13 percent of the sample) said they chose the house for its "convenient location;" 21 households (about five percent) wanted to be "closer to work;" and 11 (about three percent) wanted to be "closer to family." Also, 32 households (about eight percent) cited "amenities of the neighborhood," although the specific nature of these amenities (ex. availability of cable TV or corner drug store) is unclear. This suggests that accessibility measures which incorporate many types of activities in which a

household would participate might have more explanatory power than a measure which relates accessibility to only one type of activity, such as work.

Butler (1977) shows that accessibility in terms of distance to the CBD — a suspect instrument for the true access to opportunities at best — is necessary as an explanatory variable to avoid bias in estimates of hedonic prices in rental markets. He found that the measure was not significant in determining the price of owner-occupied housing. However, a more realistic measure of accessibility might prove significant in explaining location choice in both rental and owner-occupied housing markets.

Chapter 3

Theoretical and Analytical Issues

Before we further examine accessibility, and its relation to the transportation and land use systems, it will be useful to further clarify what we hope to measure as "accessibility," and how this choice affects our data needs and calculation methods. Some questions which must be answered include whether or not accessibility should be determined in part by actual travel patterns, to what levels we can aggregate individual trip decisions without destroying information from the variation among individuals, what determinants of accessibility are realistically available, and how the limited capacity of the transportation network constrains accessibility. Once we have decided on a reasonable framework for measuring accessibility, we should address how we expect urban dwellers to respond to changes in accessibility. Finally, it may also be useful to ask how accessibility can be used to compare the quality of life in different urban areas, or to explain how cities compete in a regional and national context.

3.1. Actual or Potential Travel

Morris, Dumble and Wigan (1979) present both viewpoints on the issue of whether accessibility should reflect actual or potential travel. Part of this problem arises from the difficulty that traditional accessibility measures have in summing the attractiveness of various destinations. For example, the

Hansen-type measures based on the gravity model assume that accessibility decays exponentially with travel time if attractiveness is constant. Lerman's model (1975) uses a decay function based on a random utility model which considers only travel time. Calibrating such decay functions would require knowledge about current trip patterns, and might therefore prejudice accessibility measures by giving extra weight to existing travel patterns, and not being responsive to changes in the transportation or land use networks.

The argument for attempting to measure *potential* travel is based on the premise that choices have value, even if they aren't selected. As an analogy, consider the popularity of "super" grocery markets over locally owned neighborhood grocery stores. Obviously, a customer at a super store doesn't intend to purchase one of each product or brand. Rather, the increased product variety at a super store allows a customer to choose a brand which might yield more satisfaction than those brands available at the corner market. Likewise, that customer would have the opportunity to trade off cost with desirable product attributes, possibly taking advantage of weekly specials. A customer with a choice between two brands is better off than one who has no choice, even if that single "option" is the brand the customer would prefer if he or she had a choice. Similarly, trip makers who have choices of multiple destinations at which to conduct their activities are better off than those with more restricted choices. Proponents of measuring potential travel, including the author, argue that accessibility measures should attempt to reflect the value of these unselected alternatives.

By considering travel costs and destination attractiveness together, we can avoid some of the arbitrary criteria used in traditional accessibility measures which are based on either actual travel or potential travel. Travelers don't ask "at what rate do destinations look less attractive?" or "is the destination within X minutes?" but instead "is this particular trip worth it?" In other words, does the utility gained at a destination outweigh the disutility incurred in traveling to that destination? By summing the net utility possible at all destinations, we have an idea of the accessibility of a given location.

A similar issue is how to address accessibility via different travel modes. Should separate accessibility measures be calculated for each mode, greatly increasing the complexity of the data processing involved? Or if mode-specific accessibility measures are to be combined, should they be weighted by observed mode shares, by mode shares predicted by a behavioral model, or by some other method? Is there a way to reflect the added flexibility of choosing alternate travel modes without double-counting?

3.2. Appropriate Level of Aggregation

In the most general sense, accessibility measures are an aggregation of potential trips — over destinations, modes, people, and origins. This section has hinted at aggregation over destinations and modes. Aggregation over locations and trip-makers is discussed below.

3.2.1. Decision-making units. Ideally, we would like to consider the accessibility of each decision-maker individually. One obvious constraint to this approach is the sheer computational capacity required to model the behavior of say 4 million people in a metropolitan area, such as Boston. Another complication pointed out by Chapin (1974) is that individuals behave in a hierarchical fashion. For example, a family may make a single decision on where to live, and what types of trips are made by its members. In this instance, the family acts as the decision-making unit, although it considers the wants and needs of its members. Group living arrangements add even more complication. Should two people who met via the classified ads to rent an apartment be considered as two individuals or a single household? College dormitories, sororities, and fraternities, which often have various committees and circles of friends, would have more decision-making levels.

A similar difficulty occurs on the neighborhood level. Although a central neighborhood decision-making organization may not exist, one might expect the travel patterns of members of the same neighborhood to be similar. Transportation planners have traditionally grouped people in to geographical and political "Transportation Analysis Zones," and considered the average individual within each zone. However, there are several drawbacks to this approach. Residents form their own natural association of who and what constitutes their neighborhood, which may be completely unrelated to the lines transportation planners and the Census Bureau find convenient to draw on a map. Furthermore, individuals within a

neighborhood are different; not only is it useful to know the average value of a particular characteristic of a neighborhood, but we also want to know the variety of that characteristic throughout the neighborhood. The statistical concept of variance is especially useful with scalar attributes such as income, but harder to formalize with categorical variables such as race.

3.2.2. Decision time frames. The specific definition of an accessibility measure implicitly assumes a time frame within which trip decisions are being made. The more specific a measure, the shorter run decision being considered. If we define "short run" to mean individual trips, a useful accessibility measure would consider the specific purpose of the trip. Clearly, such an objective would change from trip to trip, creating a necessity for a multitude of accessibility indices. Since we are interested in more long run decisions involving the location of residences and firms, using a more general accessibility measure may be a better approach. Such a measure would incorporate all types of trips which an actor is likely to make, and the effect of limited information available when the location decision is made. However, we should realize that not all location decisions (or transportation investment decisions for that matter) are made all at once, so the intermediate run may be the most appropriate, since changes in location may be traded off with greater dedication of resources to transportation. Our choice of time frame will also be influenced by the time frame measured by the data we use to describe accessibility.

3.3. Data Limitations

Because the type and quality of data available affect how closely we can approximate our ideal accessibility measure, it is important to discuss what limitations we are likely to encounter, and the likely result of these limitations. We first examine some traditional biases in the transportation planning field — those of considering only mean values and of looking at a city only in plan view. The we question whether "more is better;" will extensive, detailed geographic information help us better explain location decisions in the context of accessibility?

3.3.1. Variety within population. The traditional four-step transportation planning process conducts analysis by assuming that all individuals in a "zone" are identical, or that little information is lost by considering only the average individual in a zone. The result is forecasts which appear to have a great deal of precision, but deceptively overlook inherent errors. (Four-step model which have errors of ten percent of less are considered "very good," and errors of forty percent are not unheard of.) Certainly some error is unavoidable, because travel surveys and censuses cannot be conducted continuously. However, by considering some measure of variance along with the mean, we may be able to significantly reduce some sources of error. (The mean and variance don't completely characterize the distribution of a variable, but the gain from additional statistics is usually small.) The census and other data sources sometimes report variance or a similar measure of

spread, except when the limited size of an area makes it impossible to preserve individuals' privacy.

3.3.2. "Horizontalness" of geographic information. The widespread use of geographic information systems (GIS) means that some very detailed information (such as street lengths) is available to transportation planners. However, this information is generally limited to a horizontal plane. In cities such as Chicago, mixed-use zoning permits apartments in buildings above stores at ground level. To a person in one of these apartments, the shop on the ground floor isn't as accessible as the TV remote control, even if the horizontal separation of both is zero. A similar phenomenon is office skyscrapers which have services such as banks or cafeterias located in their lobbies or lower floors. On the scale of a metropolitan area, vertical distances are negligible, but issues of development density and mixed-use zoning are often important to urban planners. Also, when horizontal aggregation areas are larger, this problem reduces to simply another part of the heterogeneity problem.

3.3.3. Desire for detailed economic variables. A naive approach to explaining accessibility and its affect on transportation and land use is to collect information about an inordinate number of explanatory variables. Clearly there is a point where the information gained by adding variables does not justify the expense of data collection. Even if transportation planners could obtain data at minimal cost, it is unrealistic to assume that city residents would also be able to obtain as extensive a database, much less make location

decisions based on that information. Wallace and Sherret (1973) remind us that objective engineering or financial data are only instruments for the perceptions of actors in a city, and such instruments should be used cautiously.

3.4. Capacity of the Transportation Network

Until now, we have considered only demands placed on the transportation network. In any economic model, including a model of accessibility, supply is as important a factor as demand. For transportation, supply is traditionally expressed in terms of capacity. Capacity constraints may be modeled as part of a simultaneous equilibrium (see Anas, 1985, for an example) or as part of an iterative feedback process. Complications arise because the capacity of a transportation *network* is not simply the sum or minimum of the capacities of the network's components or links. How capacity affects two classes of actors is examined below.

3.4.1. The household's perspective. Throughout this document, we have largely considered the household's ability to consume. This analysis has considered two important components of consumption: the benefits of activities at locations throughout a city, and the costs associated with traveling to those activities. Capacity therefore affects the transportation costs or disutility. Highway traffic modelers are familiar with congestion (the exhaustion of capacity) causing a degradation of travel times. The response of transit to use is more subtle. Increased ridership affects travel time since

longer dwell times at stops are necessary to allow passengers to board and alight. Perhaps more importantly, ridership directly affects the load factor of a transit vehicle, which impacts passenger perceptions of comfort.

An important consideration arises here. We have defined accessibility in terms of *potential* travel. Since travel disutility is a function of both capacity and use, we need to know *actual* travel as well. Predicting future actual travel with sufficient accuracy is indeed difficult. One way around this dilemma may be to check that the level of congestion is consistent with our estimates of accessibility. Another way of expressing this notion in our stochastic framework is to consider a simultaneous equilibrium of the *expected values* of travel and congestion.

3.4.2. The business's viewpoint. Firms are considered separately because their accessibility objectives are different from those of individuals. Firms "consume" labor as an input, and need to get their products to a market in order to make a profit. Since individuals are both suppliers of labor and purchasers of products, a firm will view its attractiveness (or accessibility to labor or customers) in terms of those individuals' travel disutility, which is affected by capacity. In addition, capacity represents an upper bound on a firm's potential markets.

3.5. Affect of Improved Accessibility

Knowing the level of accessibility of a single transportation and land use system is generally only useful for evaluating equity among groups of urban residents. However, a consumption-based accessibility measure holds much more promise to transportation planners — the ability to evaluate different urban investment programs. In order to conduct such an analysis, one must know how travelers respond to increases and decreases in accessibility. It is useful to consider traveler responses in the economic terms of income and substitution effects, which traditionally describe consumers' responses to changes in product prices. Since accessibility incorporates both price and quantity information, this division seems reasonable.

3.5.1. Income effect: more travel, greater opportunities. The income effect can be summarized as a fall in the price of one good allowing increased consumption of all goods. In the case of improved accessibility, fewer resources (time, money and effort) need to be dedicated to a fixed amount of travel, so residents will both travel more (in terms of number, and length or disutility of trips), but they will also participate in a greater amount of activities (that is, they will consume more). This is similar to a result of transportation investment argued by David Aschauer (1989) on a national level: increased transportation investment leads to greater productivity and improved competitiveness.

3.5.2. Substitution effect: competing neighborhoods. The substitution effect can sometimes be thought of as working counter to the income effect. When the price of one good falls, consumers will spend a greater *proportion* of their income on that good, and less on other goods. When access improves, residents may partake of activities in neighborhoods which are most accessible to them at the expense of other neighborhoods. The substitution effect may lead to subdivisions of a metropolitan area needlessly fighting over geographically limited benefits. Therefore, it is important to know the relative sizes of the substitution and income effects, in order to know if increased accessibility would actually be healthy to a region as a whole.

3.6. Extrapolation to the Intercity Level

Since accessibility is a useful concept for comparing the distribution of transportation and activity within a city, it seems logical to ask if cities can be compared to each other.

3.6.1. Cross-city comparisons. If accessibility indices are easily calculated for neighborhoods within one metropolitan area, with similar information for another metropolitan area, it would seem to be easy to determine if one city is more accessible than another. However, planners must be very careful when attempting to borrow data from other cities. Residents of the two urban areas may value travel and location attributes differently, which would in fact influence their decision of which city they prefer to live in. If the implied valuation of attributes is estimated for only one metropolitan area, using

those values to calculate accessibility outside of that area would be meaningless. Cross-city comparisons would require data from all cities to be compared and a more careful specification of the valuation of various attributes in order to avoid the self-selection bias.

3.6.2. Accessibility to productive resources. Another use of the accessibility framework is simply to consider economic decision makers at one level up: the relation of cities to each other. Cities don't exist in isolation, but trade with each other. Often, cities specialize in certain industries such as financial markets, electronics technology, or agriculture. Similarly, tourism in different cities are not perfect substitutes. In such a framework, the federal government might evaluate how a high-speed rail network (to use a current example) affects national productivity and competitiveness through cities' improved access to productive resources and goods markets.

Chapter 4

Model Development

This model of accessibility takes a microeconomic consumption approach. Throughout this discussion, "consumption" will refer to purchases of every imaginable type of good, service, or opportunity. Some of these purchases might be quite necessary to life, such as buying groceries, health care, and clothing. Other more frivolous or hedonistic purchases, such as recreation or status goods, may better fit the typical connotation of consumption. We are not interested in making value judgments on the type of goods and services that people buy, but rather in how people's needs and desires influence their use of the transportation network and the various establishments within a city.

Since accessibility represents potential travel and consumption, no budget constraints on income or time are imposed. However, we do impose a "rationality" constraint that no trips are made unless the gains from consumption at the destination offset the difficulty of travel there. The reality of incomplete and imperfect information leads us to adopt a random utility model.

4.1. Simplifying Assumptions

Several assumptions have been made to simplify the analysis that follows. These are discussed below.

4.1.1. Additive utility of attributes. A standard assumption of economic models is a simple additive utility function describing consumers' preferences for goods and services. This leads to the result that as consumers' incomes increase, they purchase a greater quantity of the goods in the same proportion as they did before. Modeling utility in terms of attributes rather than products allows for slightly more realism since the availability of attributes is determined by their mix in products. Also, some of the income effect on consumption patterns will be captured by estimating different accessibility indices for different groups, based on income and other socioeconomic factors. However, in order to fully model satiation, as well as inferior and luxury goods, we would need a polynomial or piece-wise-linear utility specification. The drawbacks of this latter approach are more parameters to estimate, and polynomial or piece-wise-linear combinations of right-hand-side variables would be highly correlated, causing an efficiency loss in parameter estimates. For each socioeconomic group considered, we model utility as a simple, linear combination of attributes, and neglect satiation.

4.1.2. Trip chaining. In the short run, a traveler might be aware of several needs that would have to be fulfilled within roughly the same time frame.

The traveler might then combine several trips to satisfy these needs into a single trip chain or tour. However, since needs vary over time and in frequency, it is not possible to know beforehand what needs may occur simultaneously. We assume that travelers do not consider the possibility of chaining trips in the long run when decisions such as location choice are made.

4.1.3. Hedonic utility values consistent between short and long runs. Since our model is one of land use decisions, we are implicitly modeling a long run response. However, some of the proposed uses of this accessibility index, namely using accessibility as a better instrument than the change in user costs for the social benefits of a transportation program, assume benefits accruing from short run trip decisions. In order for accessibility to be used for this purpose, some stability between the long run and short run should exist.

4.1.4. Joint distributions of stochastic disturbances. Although this assumption is inessential, we assume that the stochastic disturbances or combinations of disturbances of utility components are independent and identically distributed Type 1 Extreme Value random variables in order to utilize nested logit theory. Ben-Akiva and Lerman (1985) provide a concise reference.

The i.i.d. assumption is restrictive for two reasons: (1) the variance of the disturbance is assumed to be constant for each mode — for example, that walking has as much inherent randomness as driving or riding a bus; and (2) modes such as bicycles, buses, and autos travel on the same roadway, and

would all experience the same random disturbance related to the functioning of that roadway, so these modes' error terms might not be independent. However, dropping the i.i.d. assumption would require much more complex and computationally intensive estimation procedures, such as multinomial probit, so we retain the assumption for simplicity.

4.2. System Specification

In making long run location decisions, residents and firms consider accessibility along with characteristics of the areas where they might locate. The distribution of activities throughout an urban area affects the day-to-day and minute-to-minute travel demands. However, over time, a pattern of travel demand should develop. Travel demands are constrained by the capacity of the transportation network, and thus affect the level of service available from the network. Political decision makers will consider both existing land use patterns and the service levels of the transportation network when considering investments that improve the functioning of transportation and other infrastructure. The relations among these decisions and their results are shown schematically in Figure 4-1. Transportation investment decisions are treated exogenously due to the highly political nature of these decisions. This research intends to make a contribution in improving the accessibility model. Other modules in the system represent the current use in practice and are described for completeness only. (For example, planners wishing to make forecasts of residence choice might wish to use a model

Figure 4-1. System Schematic
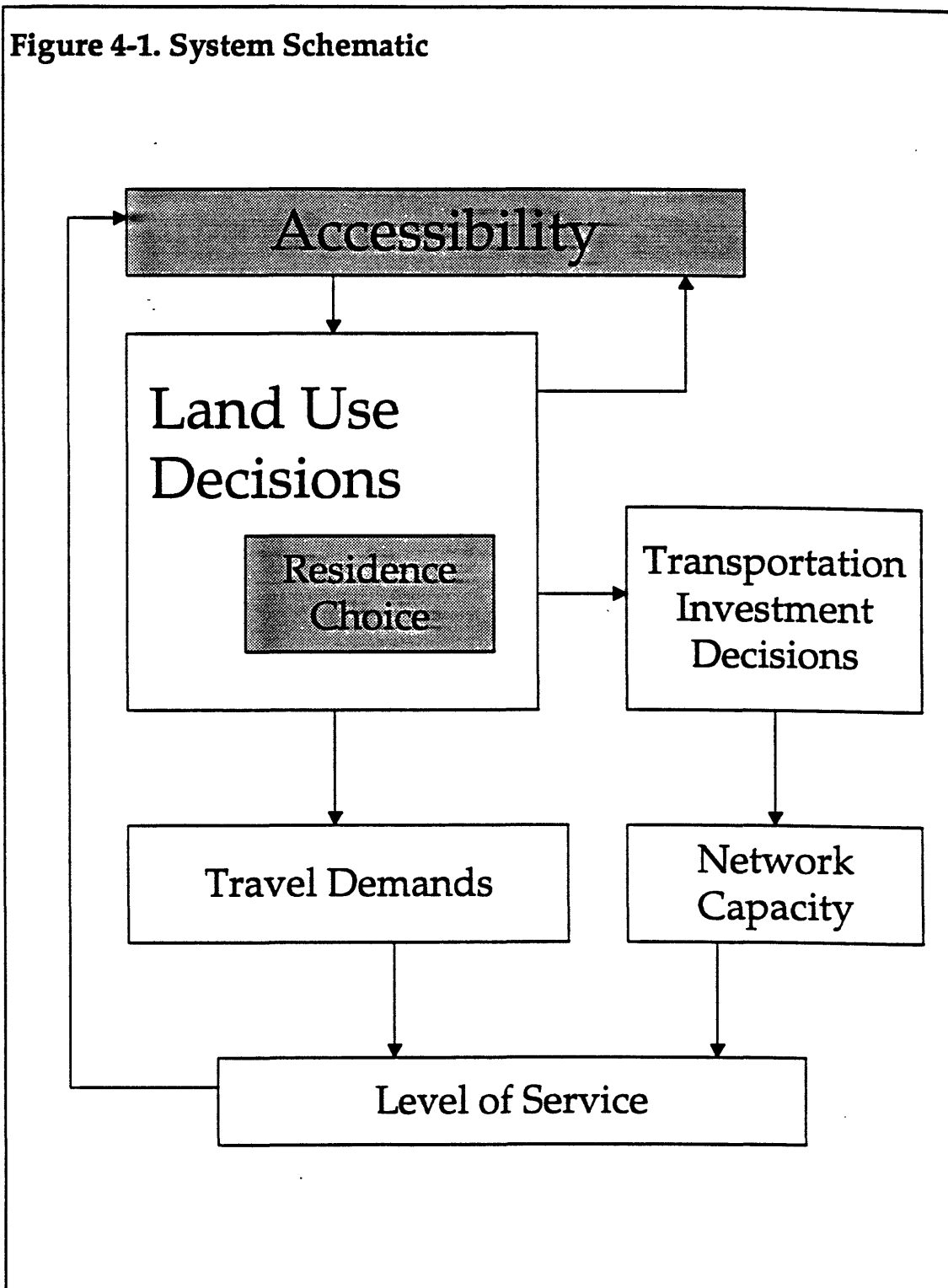
similar to one in Brand, Barber and Jacobs, 1967; Lerman, 1975; or Weisbrod, 1978.) We will concentrate on examining the accessibility of residents, but not firms, and therefore be interested in housing markets.

There are two convergent descriptions for the derivation of the accessibility model, both of which are presented below. One process is using economic theory and inductive reasoning to first consider individual trips, and then to expand this concept to the entire geographic area in which a person may make trips. The second derivation is based on a hierarchical model of travel and residence choice.

The inductive approach began with questions about what considerations are important to people when they travel, and how do these people then strategically choose places to live based on their expected travel patterns. This led to the conclusion that what is important is "what you can get to" and then to questions about what kind of concepts accessibility should and should not measure. The definition of accessibility was refined and stated mathematically, borrowing from microeconomic and discrete choice theory.

However, working within the framework of discrete choice theory — primarily sequential choice theory — will lead to an identical accessibility measure. It is common practice for nested choice models to include "inclusive values," which reflect the benefit of being able to choose among several alternatives at a "lower" level in the choice model. That is, making one type of choice, say where to live, allows one to make further choices that are

conditional on the first decision. However, the possible outcomes of the second set of choices, which aren't known ahead of time, influence the attractiveness of alternatives within the first choice, so the inclusive value serves as an *expected* outcome of the later choices.

One might wonder why a sequential or nested choice structure arises at all. The answer is that nesting is a way to handle correlation among unobserved (by either the analyst or the economic actor) attributes of alternatives. This correlation may arise for several reasons. One is a difference in the frequency that different types of decisions are made. For example, residence decisions may be made only once a year, but travel decisions can be made every day. Differences in decision frequencies often arise from long term contracts (such as a 12-month apartment lease) or high transaction costs (such as the "closing costs" associated with the purchase of a home). Other natural or physical phenomena may cause attributes of alternatives to be correlated. For example, many travel paths exist to the same destination, and several of these paths might have common segments on the same street, with the same scenery. The unobserved attributes of the common street segments and of the final destination would cause path choice to be viewed as a subchoice of destination choice.

Also, it is important to remember that nesting and aggregation are not necessarily synonymous. A nested model structure is appropriate when alternatives may be correlated through unobserved attributes. In contrast, aggregation may occur merely for government convenience and does not

necessarily imply correlated attributes. For example, a town or ZIP code boundary may encompass households that do not identify with that geographical unit. Instead, a household may identify with some concept of a neighborhood that has vaguely defined boundaries or straddles readily identifiable geopolitical boundaries.

4.3. Accessibility Model

We now proceed to define a specific model of accessibility based on the concerns and assumptions mentioned above.

4.3.1. Components of accessibility. We define accessibility in terms of the utility possible from consumption at locations throughout an urban area, net of the disutility of the travel required to reach those consumption possibilities. In the derivation of an accessibility index, it is useful to consider the behavioral relation between *individual* residences as origins and *individual* establishments as destinations. Accessibility to all activities in a region can be thought of as an appropriate aggregation of a similar notion of access between a single origin and a single destination. This seems a natural approach as it allows us to better explore and understand the components of access utility. Later, we will want to aggregate the access of origin-destination pairs in such a way that destinations with consumption possibilities which don't offset the travel required to reach them will not contribute to the accessibility of a location. If, for the moment, we call this aggregation function h(•), we can write

$$Accessibility_i^g = h\left(ijAccess_{ij}^g, j \in C_j\right) \qquad\qquad (4\text{-}1)$$

where i represents a given origin residence, j represents a destination establishment, C_j represents the set of all destinations, g represents the socioeconomic group under consideration (for now, we can assume that individuals form a socioeconomic group of one, or that they belong to readily identifiable, perfectly homogenous socioeconomic groups), $Accessibility_i^g$ represents the accessibility of origin i to all destinations for socioeconomic group g, and $ijAccess_{ij}^g$ represents the accessibility of origin i to destination j for group g which we will detail below.

Access utility for a single origin-destination pair consists of two components — the utility gained by consumption opportunities at the destination (which is a characteristic of that destination, but not influenced by the origin), and the disutility of traveling from the origin to that destination.

$$ijAccess_{ij}^g = DestinationUtility_j^g + TravelDisutility_{ij}^g \qquad\qquad (4\text{-}2)$$

In his model of shopping mall trip distribution, Wheaton (1993) calls this sum "Net Shopper's Utility," but does not extend this concept to the broader notion of accessibility. Ingram (1972) calls this variable "relative accessibility," and uses the term "integral accessibility" for what we are describing as simply $Accessibility_i^g$. We prefer to explicitly state when we are

using the generic term for describing accessibility to only a single destination, and to reserve the term "relative accessibility" for a normalized accessibility index (for example, the accessibility of a suburban area divided by that of a central city area selected as the standard of accessibility).

4.3.2. Destination utility model. The utility of consumption at a destination depends on attributes of that destination, such as the presence of a particular product. The size of a shop, crime rate, and presence of trees are also legitimate destination attributes (some of which are easier to measure than others). A linear model of utility may be written as

$$DestinationUtility_j^g = D_j v^g + \varepsilon_1 \tag{4-3}$$

where D_j is a (row) vector of attributes of destination j, and v^g is a (column) vector of values that each member of group g associates with each destination attribute. The utility value of an activity, v^g, is assumed to be net of the time and money costs associated with that activity. The disturbance term ε_1 reminds us that destination utility is a random variable, since individuals may have imperfect information or taste differences within groups, and a product may have uncertain availability at a destination.

4.3.3. Travel disutility model for multiple modes. Calculating travel disutility is somewhat less straightforward because of the many modes and paths available for travel between two points. Let m index a mode and path combination. It is typical in discrete choice models to write

$$TravelDisutility_{ijm}^g = L_{ijm}\beta^g + \varepsilon_m \qquad (4\text{-}4)$$

where level-of-service variables in L_{ijm} include in-vehicle time, walk time, wait time, out-of-pocket cost, and comfort. β_k^g is the value a member of group g associates with each level of service variable, and like travel disutility, has an implied negative sign. (That is, the coefficients on level-of-service variables that make travel more onerous, such as time, have negative signs. Variables that make travel less onerous, such as comfort, will have positive signs.) Modes that should be considered include auto, transit, walking, and bicycle. To insure a consistent treatment of access utility, we will define accessibility in terms of the activity involved with one trip, which means that level of service variables should reflect round-trip, and not one-way, quantities. (Future researchers expanding this framework to incorporate the short-run phenomenon of trip chaining may wish to use one-way level of service variables; in this case, values of the travel disutility or destination utility parameters are adjusted by an appropriate factor.)

Network performance is one of many factors giving rise to the random term ε_m. The assumption of rational behavior means that travelers will select the mode with the highest (i.e., least negative) utility. Domencich and McFadden (1975), Ben-Akiva and Lerman (1985), and others use the property that if independent random variables $U_1 ... U_n$ are Type 1 Extreme Value distributed with common variance σ and expected values $V_1 ... V_n$, then $\max(U_1 ... U_n)$ is

Type 1 Extreme Value distributed with variance σ and expected value

$\ln \sum_{i=1}^{n} \exp(V_i)$. The operator $\ln \Sigma \exp (\bullet)$ is frequently referred to as "logsum."

At first glance, the logsum operator may not seem intuitive. It isn't immediately grasped the way a simple mean or a weighted average might be. Such a formulation may be summarily rejected by analysts who do not understand how it arose, so it is useful to examine logsum more thoroughly. When constructing an inclusive value, analysts may often choose to use a weighted average of utilities from the sub choice. However, this construction would contradict our economic model that says people choose the *best* alternative, not merely an *average* one. Since utilities of the sub choice (and all choices) are assumed to be random, an inclusive value should therefore be the *expected value* of the *maximum* of these utilities. We have just shown that when the stochastic components of the utilities are i.i.d. Type 1 Extreme Value, the expected value of the maximum can be calculated with the logsum operator. For the sake of giving the reader confidence in using logsum, the functional is graphed with the maximum operator in Figure 4-2. The figure can be considered to represent the simplest case of two alternatives. One alternative can therefore be arbitrarily defined as having no (net) utility. When one alternative is clearly better than the other (X is strongly positive or negative), the two operators produce the same arithmetic result. However, when the alternatives are "close," logsum gives a greater numerical value. This is exactly the property we desire, since urban dwellers would get

Figure 4.2. Comparison of Max and Logsum Operators
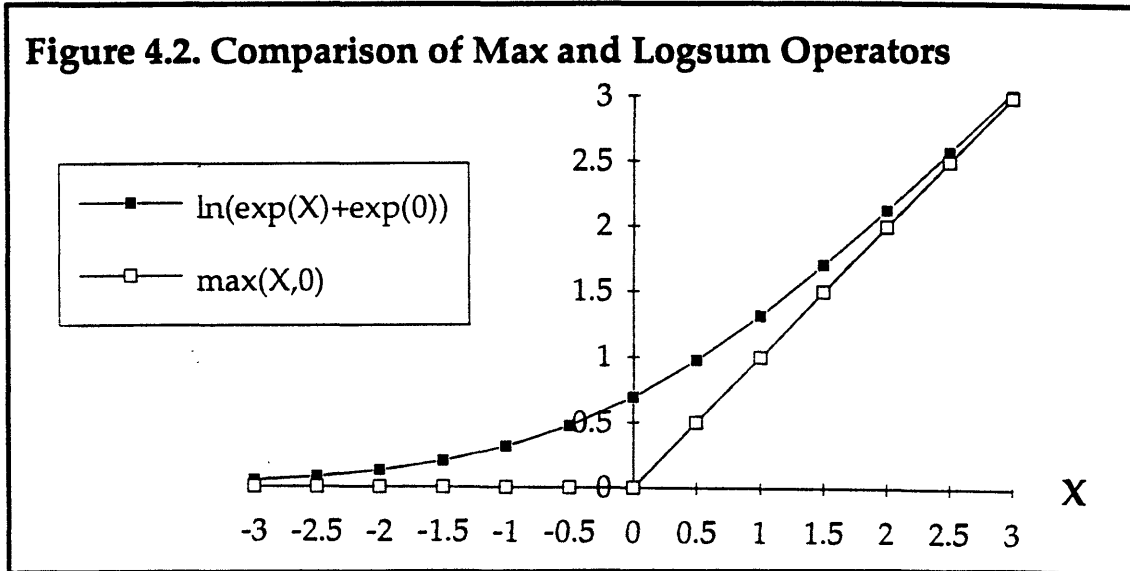
greater benefits from having two near similar choices, than having the same best choice but a much inferior second choice. The "smoothness" of the logsum function can also be viewed as arising from the uncertainty over which of two similar alternatives is actually "better."

Applying logsum to the modal disutilities, we obtain

$$TravelDisutility_{ij}^g = \ln \sum_m \exp(L_{ijm}\beta^g) + \varepsilon_2. \tag{4-5}$$

Substituting equations 4-3 and 4-5 into equation 4-2, we have

$$ijAccess_{ij}^g = D_j v^g + \rho^g \ln \sum_m \exp(L_{ijm}\beta^g) + \varepsilon_3 \tag{4-6}$$

where $\varepsilon_3 = \varepsilon_1 + \varepsilon_2$, and $\rho^g = \left(\dfrac{Var(\varepsilon_2)}{Var(\varepsilon_1) + Var(\varepsilon_2)}\right)^{\frac{1}{2}}$.

4.3.4. Correction for using aggregate destination data. Transportation planners generally aren't lucky enough to have information about every individual establishment. Instead, data is often aggregated over geographic areas and stratified by ranges of variables of interest. Following the discussion in Ben-Akiva and Lerman (1985), the utility for an "aggregate alternative" (such as the multiple destinations in a zone) must be corrected for a "size effect" and a "heterogeneity effect." Let J index destination zones (or census blocks) composed of multiple j's, and V_J and V_j stand for the systematic utilities of aggregate (zone) and elemental (establishment) alternatives. Similarly, j and J can index destination attributes $D_{.}$ and level-of-service variables $L_{i \bullet m}$.

The aggregation correction is

$$V_J = \underset{j \in J}{E}\left[V_j\right] + \frac{1}{\mu}\ln B_J + \frac{1}{\mu}\ln M_J \qquad (4\text{-}7)$$

where M_J is a measure of the size of J (in terms of the number of elemental j);

$B_J = \dfrac{1}{M_J}\sum_{j \in J}\exp\left(\mu\left(V_j - E\left[V_j\right]\right)\right)$, which measures the heterogeneity in J; and μ

is related to the variance of the stochastic terms of the elemental alternatives. Often B_J and M_J are unobservable. The analyst may not even be certain what

constitutes an elemental alternative. (For example, do travelers view the individual store as a single destination, or a certain shopping district?) However, we can usually acquire instruments for M_J such as the area or number of establishments in a zone. Therefore we can approximate this correction using

$$M_J = S_{DJ}\gamma_D + \varepsilon_4, \tag{4-8}$$

where S_{DJ} is a row vector of destination size variables. For lack of instruments, it is typical to neglect the B_J correction and assume a sufficient degree of homogeneity within an aggregate alternative. However, if different variables are available at different levels of geography, it might be possible to use information from the variable with the finer geography to construct an instrument for B_J. For example, suppose that retail employment density is available on only the town level, but the density of grocery stores is available on the census block level. Such a situation would be typical since stores like to advertise where they are located, but prefer to keep staff records confidential. The two average densities for the town level would be used to construct D_J, and the additional information about grocery store density would be used as an instrument for B_J as follows.

Let $D_b = \underset{j \in b}{E}[D_j]$ for every element D_j of D_J, where b represents the intermediate level of aggregation (b is a mnemonic for block group). Also consider the summation and division by M_J in the formula for B_J to act as an expected value operator. A Taylor series expansion of B_J gives

$$B_j = E\left[1+\mu\left(V_j - E[V_j]\right)+\frac{1}{2}\mu^2\left(V_j - E[V_j]\right)^2 + \text{higher order terms}\right].$$

By definition, $E\left[V_j - E[V_j]\right] = 0$ and $E\left[\left(V_j - E[V_j]\right)^2\right] = Var(V_j)$, so by neglecting the higher order terms (which correspond to the skewness, kurtosis, etc., in the systematic utilities due to variety in destination attributes) we have

$$B_j \approx 1+\frac{1}{2}\mu^2 Var(V_j). \tag{4-9}$$

Since $V_j = D_j v$, $Var(V_j) = v'Var(D_j)v$. However, we know $\underset{b \in J}{Var}(D_b)$ rather than $Var(D_j)$. Since $D_b = \underset{j \in b}{E}[D_j]$, we can shown that

$$\underset{j \in J}{Var}(D_j) = \underset{b \in J}{Var}(D_b) + \sum_{b \in J}\underset{j \in b}{Var}(D_j).$$

We still have no information about the within-subgroup variation, but the between-subgroup variance provides a better instrument for the within-group variation than nothing at all. Our instrument for the heterogeneity effect is therefore

$$B_j = 1+\frac{1}{2}\mu^2 v'Var(D_b)v. \tag{4-10}$$

If we define iJAccess as the accessibility of a given place to an aggregate destination J, we obtain

$$iJAccess_{ij}^{g} = D_{j}v^{g} + \rho^{g}\ln\sum_{m}\exp\left(L_{iJm}\beta^{g}\right) + \frac{1}{\mu_{1}}\ln B_{J} + \frac{1}{\mu_{1}}\ln M_{J} + \varepsilon_{5}. \qquad (4\text{-}11)$$

The parameter μ_{1} is a function of the unexplained correlation between elemental alternatives:

$$Corr\left(\varepsilon_{3j}, \varepsilon_{3j'}\right) = 1 - \left(\frac{1}{\mu_{1}}\right)^{2}, \text{ so}$$

$$\frac{1}{\mu_{1}} = \sqrt{1 - Corr\left(\varepsilon_{3j}, \varepsilon_{3j'}\right)}.$$

The parameter can also be though of as a "nesting coefficient" (coefficient on an inclusive value) in a model where elemental alternatives are a sub choice within an aggregate alternative. Therefore if μ_{1} is estimated to be near one, the remaining parameters in equation 4-11 are not sensitive to the aggregation scheme used to classify destinations into zones. That is, when μ_{1} is exactly one, there is no correlation among unobserved attributes of elemental alternatives, and a non-nested multinomial logit model performs as well as a nested logit model. If a data generating process is multinomial logit, but the analyst specifies a nested logit model, there is no misspecification error since multinomial logit is a special case of nested logit.

(However, the reverse case of misspecifying a nested model as a multinomial logit model results in the famous "Red Bus-Blue Bus" paradox.)

4.3.5. Aggregation of accessibility over all destinations. Now we return to the issue of the summation or aggregation function h($\bullet$) from equation 4-1. A naive approach might be simple summation; however, we have not insured that the access utility between each origin-destination pair will be positive — that is, that the trip will be worthwhile to the person making it. We want an operator that does not penalize accessibility when some destinations aren't meaningful, but does reflect the benefits of having meaningful second and third choices.

The maximum operator obviously would not penalize accessibility when some destinations aren't meaningful, because it ignores all but the best one. Other "order statistics" (for example, fifth best) have similar properties. However, because we are maximizing over *random* variables and taking the expected value, it is not quite correct to say that inferior alternatives are *completely* ignored. If one alternative is a close second, where "close" is relative to the size of the stochastic disturbance, then the expected value of the maximum of the two independent alternatives may indeed be greater than the maximum of the expected values. The curved area around X=0 in Figure 4-2 illustrates this statement (recall that the logsum operator gives the expected value of the maximum of two or more i.i.d. Type 1 Extreme Value random variables). A nested residence and destination choice model is

suggested by the notion that people choose a house based on their ability to make important trips from that origin.

One might then hypothesize that if an operator exists to express the utility of the best choice (that is, logsum), would not a similar operator exist to express the utility of the top several choices? The difficulty here is that discrete choice theory is structured around the analysis of a *single* choice. While we would like accessibility to incorporate many trips, the mathematics for doing this is not necessarily appealing. One method would be to construct a different accessibility index for every imaginable trip purpose. Obviously this would significantly increase data requirements and make estimation more difficult. Also, the result would not yield a simply method for testing outcomes of different transportation, land use, or other government (or even private sector) investment schemes. Another framework is that instead of modeling the choice of a single destination, we construct alternatives that are *packages* of destinations. Such a framework would be quite useful when used in conjunction with trip-chaining. However, questions of how many destinations to include in a package, or how to manage data for the many more alternatives are not easily answered.

Therefore, we will define accessibility as an inclusive value based on the best outcome of destination choice, since alternative specifications are not tractable. Although we have defined accessibility in terms of a single trip, our generic specification of accessibility allows that single trip to reflect the utility of any trip that a resident may take. That is, if a resident makes work,

shopping, and social trips, then employment, shopping, and social destinations would all provide utility. More importantly, the rationality constraint of "for any given trip, choose the best destination" is preserved. It might be likely that all destinations meaningful to a given traveler may have similar associated utility levels, and therefore accessibility would increase through the logsum treatment of stochastic disturbances. Substituting the logsum operator for h($\bullet$) of equation 4-1 gives that the total accessibility from an origin i is

$$Accessibility_i^g = \ln \sum_J \exp \left(\begin{array}{c} \mathbf{D}_J v^g + \rho^g \ln \sum_m \exp\left(\mathbf{L}_{iJm}\beta^g\right) \\ \\ + \dfrac{1}{\mu_1} \ln B_J + \dfrac{1}{\mu_1} \ln M_J \end{array} \right) + \varepsilon_6. \tag{4-12}$$

Estimating accessibility from a hedonic regression on rents will obviously involve nonlinear estimation. However, taking advantage of trip-making data may be useful in estimating "lower nests" of the accessibility model. The disadvantage of this second approach seems to be that of making accessibility more dependent on actual travel patterns, rather than potential travel.

4.4. Residence Choice Model

There are two processes occurring which affect location decision: (1) the competition of different land uses for the same plot of land, and (2) the

demand for all types of land use taking up the available land in an urban area. We could model the first process as land owners maximizing their profit in terms of a discrete choice between competing land uses — residential, office, retail, vacant, etc. — subject to zoning constraints. In reality, the zoning of a given parcel of land may be changed if a potential user is willing to pay certain additional costs, such as lobbying politicians and adding environmental amenities to the site design. However, to ensure tractable models, we will assume zoning is exogenously determined by political processes. We then model the choice among land used for housing, which by necessity will be occupied by different people.

Land owners will operate under the rational decision rule

maximize Profit   = Rent or Sale value

$$- (\text{land costs} + \text{maintenance} + \text{materials costs}). \qquad (4\text{-}13)$$

We assume that since the land will be used in the same manner by all potential buyers, costs will be the same to the owner. Therefore, the owner will contract with the prospective buyer offering the largest rent or sales price, so

$$MarketRent_i = \max_g \left(BidRent_i^g \right), \qquad (4\text{-}14)$$

where g represents the different individuals or groups making bids for land.

Potential land users make bids on parcels of land based on the utility they expect to derive from the use of that land. Residences gain utility from access to relevant activities, and from attributes of the housing unit and its surrounding neighborhood. Because consumers equate marginal benefits across products with the price they must pay for those benefits, bid rents will be proportional to the total utility derived from housing:

$$BidRent_i^g = \lambda_1^g Accessibility_i^g + H_i \lambda_2^g + T_i \lambda_3^g + \varepsilon_7, \qquad (4\text{-}15)$$

where H_i is a row vector of housing attributes for residence i, T_i is a vector of neighborhood ("town") variables, and the elements of λ^g are the hedonic rents that a member of group g is willing to pay for the presence of those attributes.

Like information about destination attributes, we expect the information about residence locations will be aggregated by geographical area. Let I represent a zone composed of several housing units i. We would expect that a neighborhood would be defined in a way consistent with our aggregation scheme, so $NeighborhoodAttribute_{ki} \equiv NeighborhoodAttribute_{kI}$. Likewise, since accessibility depends on level-of-service calculations between "typical" point in zones, we make no distinction between the accessibility of different points within a zone.

However, information on housing attributes may only be available as an average for the zone. It is equally likely that housing information might be classified into the frequency of units belonging to certain ranges of an attribute value. We are especially fortunate when data is cross-classified by both explained and explanatory variables. This situation is the case with our data sources; the census asks more detailed and useful questions of renters than it does of home-owners, possibly because of privacy reasons. In order to take advantage of this additional information, we model the two housing markets somewhat differently.

4.4.1. Apartment rent model. The census reports the number of rental units in a given area that have a monthly rent in one of several rent categories, and some housing attribute categories are also cross-classified by rent range. Although no aggregation method will allow us to estimate equation 4-15 directly, we will see that estimating such a model with this reporting scheme is considerably more straightforward than for other schemes.

Consider a renter responding to a census questionnaire. Essentially, that person is supplying us a series of dummy variables d_{kli} where

$$d_{kli} = \begin{cases} 1 & \text{if } b_k \leq Rent_i < b_{k+1}; \ k = 0...K-1 \\ 0 & \text{otherwise} \end{cases} \qquad (4\text{-}16)$$

and b_k represents the boundaries between rent ranges (typically $b_0 = -\infty$, $b_1 = 0$, and $b_K = \infty$), K is the number of rent ranges, I is the block group that a dwelling is located in, and i indexes the individual dwellings. (For simplicity, the superscript g referring to the socioeconomic group of the resident is omitted.) Obviously, we do not observe someone living in a unit for which $d_{0Ii} = 1$, that is, a unit with negative rent; instead we observe a vacant unit. Finally, the census reports $d_{kI} = \sum_{i \in I} d_{kIi}$ rather than individual d_{kIi}'s.

The discrete nature of the explained variable makes our model unsuitable for estimation by familiar ordinary least squares or non-linear estimation procedures. Discrete choice models familiar to transportation analysts, such as logit, are also not appropriate for this purpose because unlike the traditional application to travel modes, rent categories *do* have a natural ordering. The class of maximum likelihood estimators designed to exploit discrete, ordinal left-hand-side variables are simply called *ordered logit* or *ordered probit*, depending on the distribution of the error terms. Greene (1990) provides a description of these estimators for the general case of more than two alternatives or categories.

If we let $F(\bullet)$ represent the cumulative probability distribution of ε_7, X_{Ii} a vector of all explanatory variables $(D_J\ L_{IJm}\ B_J\ S_{DJ}\ H_i\ T_I)$, Θ^g a vector of all parameters $(\lambda^g\ \nu^g\ \rho^g\ \beta^g\ \mu_1\ \gamma_D)$, and $g(X_{Ii}, \Theta^g)$ the model in equation 4-15, then the likelihood function, $\pounds$, of observing our data set is

$$\pounds = \prod_{g} \prod_{l \in C_l} \prod_{i \in I} \prod_{k=0}^{K} \left[F\left(b_{k+1} - g(\mathbf{X}_{Ii}, \Theta^g)\right) - F\left(b_k - g(\mathbf{X}_{Ii}, \Theta^g)\right) \right]^{d_{kii}}. \tag{4-17}$$

Equation 4-17 illustrates why it is important to have our explanatory variables, particularly housing attributes, presented in a useful manner. The non-linearity of both F(•) and g(•) make it unlikely that X_{Ii} will appear as a single expected value per origin I in a reduced likelihood function. Instead, explanatory variables should be classified by rent ranges so they can appear in the correct term of the product indexed by k. Also, the aggregation of X_{Ii} into a representative X_I will produce bias unless X_{Ii} is homogenous within each I. Note that housing attributes H_i are the only variables remaining to be aggregated to the origin level. We can, however, escape the aggregation bias problem if our housing attributes happen to be discrete variables (such as the number of rooms in an apartment, or the presence of a telephone) and our data are cross-classified by that variable and the rent ranges.

The census presents rental units by number of bedrooms in almost this manner — the only complication is that units with three or more bedrooms are reported as a single category. If we assume that all units in this category have only three bedrooms, our estimate for λ_2, the hedonic price of bedrooms, will be biased upwards, compensating for our underestimate of the true number of bedrooms in that category. Likewise, estimates of other parameters that are correlated with λ_2 will also be biased.

If we define

$$d_{krli} = \begin{cases} 1 & \text{if } b_k \leq Rent_i < b_{k+1} \text{ and } Bedrooms_i = r; \\ & k = 0 \ldots K - 1; \; r = 0, 1, 2, 3 \text{ (or more)} \\ 0 & \text{otherwise} \end{cases} \qquad (4\text{-}18)$$

and $d_{krl} = \sum_{i \in I} d_{krli}$, then we can rewrite the likelihood function in terms of data available from the census, rather than individual housing attributes:

$$\pounds = \prod_{g} \prod_{l \in C_l} \prod_{k=0}^{K} \prod_{r} \left[F\!\left(b_{k+1} - g\!\left(\mathbf{X}_l, \Theta^g\right)\right) - F\!\left(b_k - g\!\left(\mathbf{X}_l, \Theta^g\right)\right) \right]^{d_{krl}}. \qquad (4\text{-}19)$$

The requirement that housing attributes be cross-classified with rent does indeed limit our choices of variables; however, we feel that the explanatory power gained by using an ordered estimation procedure with cross-classified data offsets that which might be gained by using other variables. As Chapter 5 will explain, some of the other housing variables don't capture the attribute we would like, or provide little information over our bedrooms variable.

Finally, we must address how our estimation technique handles anomalies such as vacant rental units, which may be treated differently in our data set. In the case of the housing census, limited information is available about vacant units, and an additional category of "no cash rent" is tabulated. Information about bedrooms in vacant units is reported in only one category of all vacant units, rather than for rental and owner-occupied units. We could

assume that vacancy rates are the same for both tenure types, but we have no a priori reason for doing so. Another approach would be to exclude vacant units from our data set, and calculate the likelihood function *conditional* on a unit being occupied; that is, conditional on rent being non-negative. The conditional likelihood function is

$$\pounds_C = P\left(\mathbf{X}_I, \Theta^g \middle| g(\mathbf{X}_I, \Theta^g) \geq b_1 = 0\right)$$

$$= \frac{P\left(\mathbf{X}_I \text{ observed given } \Theta^g \text{ and } g(\mathbf{X}_I, \Theta^g) \geq b_1\right)}{P\left(g(\mathbf{X}_I, \Theta^g) \geq b_1\right)}$$

$$= \prod_g \prod_{I \in C_I} \prod_{k=1}^{K} \prod_r \left[\frac{F\left(b_{k+1} - g(\mathbf{X}_I, \Theta^g)\right) - F\left(b_k - g(\mathbf{X}_I, \Theta^g)\right)}{F\left(b_1 - g(\mathbf{X}_I, \Theta^g)\right)} \right]^{d_{krI}} . \tag{4-20}$$

The second category, "no cash rent," is less problematic. A resident may give such a response when his or her landlord is a relative providing the unit, or if the resident agrees to a barter arrangement where he or she provides superintendent and emergency maintenance services in return for rent. The presence of a unit in this category gives us no information about its latent rent — such a response could occur from any monetary rent range, so by omitting observations from the "no cash rent" category, we will not be introducing any bias into our estimates.

4.4.2. Housing purchase model. Because housing attributes are not cross-classified by housing value for owner-occupied units, we take a slightly

different modeling approach for this sector of the housing market. Instead of estimating the hedonic prices of attributes of individual units, we will consider the discrete choice process of selecting a neighborhood in which to purchase a house. However, unlike the analogous case of destination choice, the objective function involved in this decision *is* observable as the value of the unit. Therefore, to take advantage of this information, it will appear that we are performing a hedonic regression on housing values, but only on the *average* housing value in an origin zone, not on the values of individual units. This will result in an efficiency loss compared to a regression based on cross-classified data being available. However, goodness-of-fit measures for this regression on average housing values may at first appear better than those for the rental market, as should be expected when using grouped data instead of individual data. (See Haitovsky, 1973, or Maddala, 1977, for a discussion of the pitfalls of comparing goodness-of-fit measures of regressions on grouped data with those of regressions on individual data.)

Using the same aggregation of alternatives procedure as before, the expected benefit of purchasing housing in a given neighborhood I is

$$E\left[Value_I^g \right] = \lambda_1^g Accessibility_I^g + H_I \lambda_2^g + T_I \lambda_3^g + \frac{1}{\mu_2} \ln B_I + \frac{1}{\mu_2} \ln N_I + \varepsilon_8 \qquad (4\text{-}21)$$

where $H_I = \underset{i \in I}{E}[H_i]$; $B_I = 1 + \frac{1}{2}\mu_2^2\lambda_2^g Var\left(\tilde{H}_I\right)\lambda_2$; and N_I, the size of neighborhood I, equals $S_{RI}\gamma_R + \varepsilon_9$. The row vector S_{RI} contains size variables such as the number of housing units and the number of households in an area. Since the

census reports median, rather than mean, housing values, we will use the median as an instrument for the mean.

Finally, substituting equations 4-8 and 4-12 into equation 4-21 gives

$$
E\left[Value_i^g\right] = \lambda_1^g \ln \sum_j \exp \left(
\begin{array}{l}
D_J v^g + \rho^g \ln \sum_m \exp\left(L_{Ijm} \beta^g\right) \\
+ \dfrac{1}{\mu_1} \ln B_J + \dfrac{1}{\mu_1} \ln\left(S_{DJ} \gamma_D\right)
\end{array}
\right).
$$

$$
+ H_I \lambda_2^g + T_I \lambda_3^g + \frac{1}{\mu_2} \ln B_I + \frac{1}{\mu_2} \ln\left(S_{RI} \gamma_R\right) + \varepsilon_{10}
\qquad (4\text{-}22)
$$

The parameters of interest for construct accessibility indices for a given population group are v, ρ, β, γ_D, and $1/\mu_1$.

4.4.3. Other housing modeling issues. The problem of using existing versus new structures can be incorporated by including construction costs in either the location utility or the land owner's profit. Since the current research focus is estimating an index of accessibility, rather than constructing a model of the housing market, I will restrict my investigation to the accessibility patterns implied by decisions of residents of existing housing stock.

The second part of the location choice model, land users' choices within the urban area, is important because we do not wish to assume that demanders of land must finally end up owning a (one and only one) parcel somewhere. "Somewhere" may end up being outside our study area. We would also desire to model homeless people as demanders of land use who don't end up

with a parcel because their bids for land are too low. This aspect of location choice seems more of a bookkeeping constraint on the first aspect of the model (bids for individual parcels).

One final caveat is that when hedonic regressions are being performed for several socioeconomic groups, including racial groups, it might be tempting to interpret estimated coefficients as evidence for or against discrimination in housing markets. King and Mieszkowski (1973) suggest that groups that have been historically discriminated against may pay more per unit of housing because of reasons unrelated to discrimination: These minorities may have larger families, which leads to more intensive use of the housing stock; or the minority may include a large proportion of immigrants who would not have as complete information about the housing market and therefore be more willing to accept higher rents. Obviously, a housing model must be carefully specified to separate these effects, only some of which are caused by discrimination. However, this research project is more concerned with examining the accessibility patterns of different socioeconomic groups, rather than discrimination in the housing market, which is why more attention has been given to non-housing components of the model. Therefore, conclusions about housing discrimination might be quite conjectural and should be avoided.

Chapter 5

Analysis and Estimation

Chapter 4 described the theory to construct a behavioral-based accessibility index, taking into consideration many of the issues raised in chapters 2 and 3. This chapter attempts to expand on the models of chapter 4 by addressing how the nature of readily available data sources affects the estimation procedure. Finally, a working accessibility model is estimated for the Boston metropolitan area.

5.1. Description of Data and Sources

Because accessibility is a convenient way of describing many characteristics of an urban area, the data required to estimate an accessibility index will necessarily come from many sources. For this research effort, we are examining accessibility patters in Essex, Middlesex, Norfolk, Plymouth and Suffolk Counties, which form the Eastern Massachusetts region.

The primary data sources are the 1990 Census of Population and Housing, files maintained by Central Transportation Planning Staff (CTPS, a support agency for the local Metropolitan Planning Organization), and summary information in the reference book *Massachusetts Municipal Profiles* (1989). CTPS combines records from the Census Bureau, the Commonwealth, and

various private sources. Information in *Massachusetts Municipal Profiles* is based on annual reports that cities and towns file with the state government.

The land-use and transportation data used in this project are unfortunately available at varying levels of geographical aggregation. Ideally we would like information about individual residences and establishments that is usually only obtainable through surveys. Geographical units used in this research are typically the cities and towns, census tracts, the "Transportation Analysis Zones" or TAZs used by CTPS, and "districts" created by the author to facilitate combining census tract level data with zonal data.

The variables used to estimate accessibility are described in detail in Table 5-1. Some simple summary statistics for many of the variables are presented in Table 5-2.

5.2. System to be Estimated

The estimation procedure is of course influenced by what explained variables are available. Accessibility, a latent variable, obviously cannot be regressed against different explanatory variables, but must be imputed as part of a larger hedonic regression of housing values. Similarly, the availability of trip tables for only motorized trips (i.e. not for walk or bicycle trips) constrains our ability to model destination choice.

Table 5-1. List of Variables

Name	Source	Description	Level of Geography
Travel Level-of-Service Variables			
ALDIST	composite	Airline distance between town centers scaled from a highway map if zones are in different towns, or $\sqrt{2TownArea}/3$ if zones are in the same town, or $\sqrt{2DistrictArea}/3$ for within the same district. (1)	district by district (2)
HDIST	CTPS	Network-generated distances for auto travel	zone by zone (3)
NHTIME	CTPS	Network-generated auto in-vehicle travel time	zone by zone
HTOLL	CTPS	Highway tolls (1987 ¢)	zone by zone
HTT	CTPS	Terminal walking times for auto travel	zone by zone
HPARK	CTPS	Hourly parking costs at destination (1963 ¢)	zone
FARE	CTPS	Network-generated fare for "best" transit path (1987 ¢)	zone by zone
NTTIME	CTPS	Network-generated transit in-vehicle-time	zone by zone
TWAIT	CTPS	Network-generated transit waiting times (calculated as one half vehicle headway)	zone by zone
NTWALK	CTPS	Network-generated walking times for transit trips	zone by zone
PNRCOST	CTPS	Network-generated daily Park-and-Ride parking cost, if appropriate (1987 ¢)	zone by zone
TDACC	CTPS	Network-generated drive to Park-and-Ride lot time, if appropriate	zone by zone

Table 5-1. List of Variables, continued

Name	Source	Description	Level of Geography
Destination Variables			
RET	CTPS	Retail employment	zone
WHOL	CTPS	Wholesale employment	zone
BOOTS	CTPS	Agricultural, forestry, mining, fishing, and construction employment	zone
MFG	CTPS	Manufacturing employment	zone
SVC	CTPS	Service employment	zone
FIRE	CTPS	Financial, insurance, and real estate employment	zone
UTIL	CTPS	Utilities, transportation, and communication employment	zone
GOVT	CTPS	Government employment	zone
RETD	composite	Retail employment density (RET/AREA)	zone
WHOLD	composite	Wholesale employment density (WHOL/AREA)	zone
BOOTSD	composite	Agricultural, forestry, mining, fishing, and construction employment density (BOOTS/AREA)	zone
MFGD	composite	Manufacturing employment density (MFG/AREA)	zone
SVCD	composite	Service employment density (SVC/AREA)	zone
FIRED	composite	Financial, insurance, and real estate employment density (FIRE/AREA)	zone
UTILD	composite	Utilities, transportation, and communication employment density (UTIL/AREA)	zone
GOVTD	composite	Government employment density (GOVT/AREA)	zone

Table 5-1. List of Variables, continued

Name	Source	Description	Level of Geography
Destination Size Variables			
AREA	Census (4)	Land area in 1/1,000 km² (or 0.1 hectare) (5)	block group
NBG	Census	Number of block groups per area	block group
Housing Variables			
BR	Census	Dwelling units tabulated by number of bedrooms	block group
MBR	Census	Dwelling units with three or more bedrooms	block group
Neighborhood Variables			
CRIME	Mass. Municipal Profiles (1989)	Crime rate in 1988 (violent or property crimes per year per 1,000 residents)	town
EDUCEXP	Mass. Municipal Profiles (1989)	Educational expenditures per pupil in 1987	town
RESPTAX	Mass. Municipal Profiles (1989)	Residential property tax per $1,000 assessed value	town
Neighborhood Size Variables			
NOOU	Census	Number of owner-occupied (such as single-family) units	block group
Housing Value Variables			
MEDHV	Census	Median value of owner-occupied housing	block group

Table 5-1. List of Variables, continued

Name	Source	Description	Level of Geography
Trip Table Variables			
TT	CTPS	Transit trips (linked)	town by town (6)
APT	CTPS	Auto person-trips	town by town (6)
PT	CTPS	Motorized person-trips (TT+APT)	town by town (6)

Notes: 1 For the derivation of the $\sqrt{2Area}/3$ formula, and a discussion of the assumptions implicit in it, see appendix A.

2 A "district" is a creation of the author, and generally represents the larger of a single Census tract or single CTPS zone (see note 2). In some cases, a district is larger, in order to have a close correspondence between the boundaries of a group of tracts and a group of zones.

3 "zone" refers to the Transportation Analysis Zones (TAZs) used by CTPS.

4 "Census" refers to the 1990 Census of Population and Housing.

5 A conversion between kilometers and miles is 1.609 km = 1 mi.

6 Within the city of Boston, the trip table is further broken down by neighborhoods, such as Roxbury or the South End.

Table 5-2. Summary Statistics

Name	Mean	Std. Dev.	Min.	Max.	No. Obs.
Travel Level-of-Service Variables					
ALDIST	2.377 mi	1.279	0	17.66	336,401 O-D pairs
NHTIME	40.18 min	19.84	1	138	617,796 O-D pairs
HTOLL	9.42 ¢ (1)	28.98	0	275	617,796
HDIST	2.28 mi	1.46	0.05	11.33	617,796
HPARK	14.26 ¢ (2)	37.99	0	221	787 zones
NFARE	222.65 ¢ (1)	133.87	0	1145	290,331 O-D pairs served by transit
NTTIME	44.36 min	23.41	0.16	146	290,331
TWAIT	20.20 min	11.50	0	60	290,331
NTWALK	16.57 min	7.23	0	36.2	290,331
PNRCOST	30.58 ¢ (1)	64.15	0	300	290,331
TDACC	3.43 min	4.41	0	39.95	290,331
Destination Variables					
RET	414.86 employees	714.98	0	12,330	708 districts
WHOL	165.95	328.67	0	2,792	708 districts
BOOTS	154.89	203.20	0	1,851	708 districts
MFG	568.65	1,272.96	0	10,357	708 districts
SVC	920.24	2,188.17	0	33,122	708 districts
FIRE	265.69	1,737.58	0	31,777	708 districts
UTIL	107.40	474.51	0	10,689	708 districts
GOVT	400.88	909.51	0	15,056	708 districts
RETD	363.2 employees per km²	2,057.0	0	44,292	579 districts
WHOLD	117.1	523.4	0	10,792	579 districts
BOOTSD	101.2	300.6	0	5,417	579 districts
MFGD	346.5	2,797.4	0	65,792	579 districts
SVCD	1,041.1	3,668.0	0	40,750	579 districts
FIRED	312.5	2,457.4	0	41,139	579 districts
UTILD	96.5	451.3	0	5,801	579 districts
GOVTD	508.8	1,549.5	0	21,583	579 districts

Table 5-2. Summary Statistics, continued

Name	Mean	Std. Dev.	Min.	Max.	No. Obs.
Destination Size Variables					
AREA	15.22 km²	150.68	0.024	3,634.3	588 districts
NBG	5.49	23.86	0	634	716 districts
Housing Variables					
MEDHV	$186,866 (3)	1,117,462	0	500,001	3,584 tracts
BR	3.041 (3)	4.266	0.897	5	923 tracts
MBR	0.068 (3)	0.747	0	1	923 tracts
Neighborhood Variables					
CRIME	28.29 crimes per year per 1,000 residents	23.69	1.45	165.38	156 towns
EDUCEXP	$3,259 per pupil	662	1,724	5,259	181 towns
RESPTAX	$10.61 per $1,000 assessed value	2.18	5.11	18.38	198 towns
Origin Size Variables					
AREA	659.7 km²	1,385.5	0.024	3,634.3	3250 tracts
NBG	102.2	225.8	0	634	3793 tracts
NOOU	206.4	168.8	0	1,409	3584 tracts

Notes: 1 Cost is presented in 1987 current cents.
     2 Cost is presented in 1963 current cents.
     3 Average is weighted by number of owner-occupied units (NOOU) to reflect regional averages.

The model structure to be estimated is presented in Figure 5-1. Except for the additional nested structure within the mode choice module, this model is identical to equation 4-12. The motorized sub mode choice module can be estimated directly using standard maximum-likelihood estimation programs for fitting logit models. The rental and housing value models, although having explanatory variables with a nested logit form, specifically our accessibility index, will require nonlinear estimation procedures such as maximum-likelihood or nonlinear least squares.

5.2.1. Motorized sub mode choice model. The variables that describe an auto trip are in-vehicle travel time, walking or terminal time, and costs which include tolls, parking, and wear and tear on the car. A transit trip would be described by fare and various time (walk, wait, and in-vehicle) components. The systematic utilities for the two motorized travel modes are

$$V_{transit} = \beta_1 + \beta_2\left(NFARE + \frac{1}{2}PNRCOST\right) + \beta_5 NTWALK$$
$$+ \beta_6 TWAIT + \beta_7 NTTIME + \beta_8 TDACC$$

and

$$V_{auto} = \beta_2 HTOLL + \beta_3 HPARK + \beta_4 HDIST + \beta_5 HTT + \beta_7 NHTIME.$$

Figure 5-1. Model System to be Estimated

Residence Choice

Housing Values

Origin   I

Accessibility =
Inclusive Value of
Destination Choice

Destination   J

Motorized
Sub Mode
Choice

Nonmotorized
Sub Mode
Choice

Auto   Transit   Walk   Bike
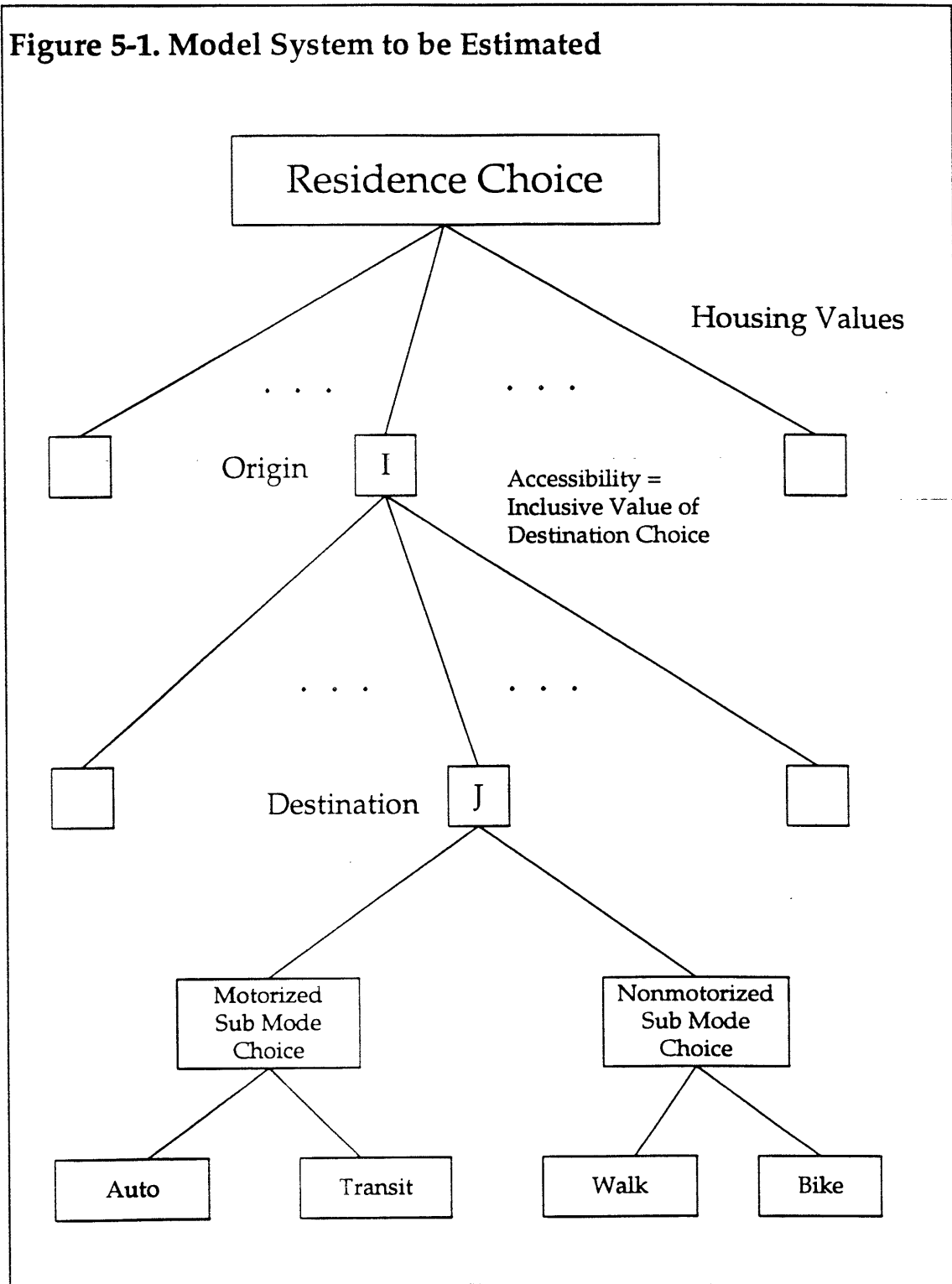
The model to be estimated is

$$P(transit) = \cfrac{1}{1+\exp\begin{pmatrix} -\beta_1 + \beta_2\left(HTOLL - NFARE - \frac{1}{2}PNRCOST\right) \\ +\beta_3 HPARK + \beta_4 HDIST + \beta_5(HTT - NTWALK) \\ -\beta_6 TWAIT + \beta_7(NHTIME - NTTIME) + \beta_8 TDACC \end{pmatrix}} . \quad (5\text{-}1)$$

The inclusive value for the utility of motorized travel is

$$IV_{motorized} = \ln\begin{pmatrix} \exp\begin{pmatrix} \beta_1 + \beta_2\left(NFARE + \frac{1}{2}PNRCOST\right) + \beta_5 NTWALK \\ +\beta_6 TWAIT + \beta_7 NTTIME + \beta_8 TDACC \end{pmatrix} \\ +\exp\begin{pmatrix} \beta_2 HTOLL + \beta_3 HPARK + \beta_4 HDIST \\ +\beta_5 HTT + \beta_7 NHTIME \end{pmatrix} \end{pmatrix} . \quad (5\text{-}2)$$

We use only transportation variables, and not user characteristics, in the mode choice models because of our intent to incorporate the mode choice models into the wider framework of accessibility. Since one of the reasons for having accessibility indices is to determine how people will react (through residence choices) to changes in transportation services, we don't want the mode choice module to assume the demographics of people living in a zone, since we later want to allow those demographics to change. This is not as restrictive an assumption as it appears. Since we may calculate different accessibility indices for different population segments, we can allow the mode choice parameters to vary across those segments.

Note that for a transit trip, we consider only a single best path. Often, particularly in studies for new capital investments for transit, planners will be concerned about the relative attractiveness of different transit paths between the same origin-destination pair. Examining the means of arrival at transit stations is a typical exercise. Appendix B discusses how the accessibility model of this document can be modified to reflect this more complex mode choice structure.

5.2.2. Nonmotorized sub mode choice model. The nonmotorized modes, walking and biking, are similar in that the traveler will likely use the same path, but will experience different travel speeds. If we let φ stand for the ratio of biking to walking speeds, we can show that modal utilities defined in terms of travel time reduce to

$$V_{walk} = \alpha_0 + \alpha_1' \, WalkTime = \alpha_0 + \alpha_1' \, \frac{Distance}{WalkSpeed} = \alpha_0 + \alpha_1 Distance, \text{ and}$$

$$V_{bike} = \alpha_1' \, BikeTime = \alpha_1' \, \frac{Distance}{BikeSpeed} = \alpha_1' \, \frac{Distance}{\varphi WalkSpeed} = \alpha_2 Distance,$$

so estimating the nonmotorized sub mode choice model with alternative-specific coefficients on distance allows us to identify φ. Since we do not have networks of walking or bike paths, we will use airline distance as the instrument for distances in this model. This assumption should be reasonable

for the shorter distance trips which are likely to utilize nonmotorized modes, although it would be a less reasonable distance instrument for trips between Quincy and Gloucester as an example. The inclusive value from the nonmotorized sub model is

$$IV_{nonmotorized} = \ln\left(e^{\alpha_0 + \alpha_1 ALDIST_{IJ}} + e^{\alpha_2 ALDIST_{IJ}}\right).$$

(5-3)

The vector of parameters α must be estimated within the appropriate housing model, since no information about nonmotorized trips is available.

5.2.3. Rental model. For the rental housing market, we estimate a hedonic rent equation based on individual data that have been censored by cross-classification and aggregation by the Census Bureau. An ordered logistic technique allows us to utilize the discrete but sequential nature of the explained variables. Since our data set includes only occupied apartments, the estimation procedure corresponds to maximizing the conditional likelihood function

$$\pounds = \prod_g \prod_{I \in C_I} \prod_{k=1}^{K} \prod_r \left[\frac{F\left(b_{k+1} - g\left(\mathbf{X}_I, \Theta^g\right)\right) - F\left(b_k - g\left(\mathbf{X}_I, \Theta^g\right)\right)}{F\left(b_1 - g\left(\mathbf{X}_I, \Theta^g\right)\right)} \right]^{d_{krI}},$$

(5-4)

or the log likelihood function

$$\ln \pounds = \sum_g \sum_{I \in C_I} \sum_{k=1}^K \sum_r d_{krI} \left(\begin{array}{l} \ln\left[F\left(b_{k+1} - g\left(\mathbf{X}_I, \Theta^g\right)\right) - F\left(b_k - g\left(\mathbf{X}_I, \Theta^g\right)\right)\right] \\ \\ - \ln F\left(b_1 - g\left(\mathbf{X}_I, \Theta^g\right)\right) \end{array} \right), \qquad (5\text{-}5)$$

where $g(\mathbf{X}_I, \Theta^g)$ is simply a compact way of writing the deterministic part of our hedonic price model

$$g(\mathbf{X}_I, \Theta^g) = \lambda_1^g \ln \sum_J \exp\left(\begin{array}{l} \mathbf{D}_J v^g + \rho^g \ln \sum_m \exp(\mathbf{L}_{IJm} \beta^g) \\ \\ + \dfrac{1}{\mu_1} \ln B_J + \dfrac{1}{\mu_1} \ln(\mathbf{S}_{DJ} \gamma_D) \end{array} \right) + \mathbf{H}_I \lambda_2^g + \mathbf{T}_I \lambda_3^g.$$

Since our rent model from Chapter 4 suggests that the error terms in our hedonic price model are i.i.d. Type 1 Extreme Value, $F(\varepsilon) = \exp(-e^{-\sigma \varepsilon})$. The variables in $g(\mathbf{X}_I, \Theta^g)$ have the same meaning as in Chapter 4, and specifically,

d_{krI} = [the number of units in rent class k with r rooms in origin block I],

$$\mathbf{D}_J = \begin{bmatrix} RETD_J & WHOLD_J & BOOTSD_J & MFGD_J & \cdots \\ \cdots & SVCD_J & FIRED_J & UTILD_J & GOVTD_J \end{bmatrix}, \text{ and}$$

$B_J = 1 + \dfrac{1}{2}\mu_1^2 v' \mathbf{Var}\left(\tilde{\mathbf{D}}_b\right) v$, where $\tilde{\mathbf{D}}_b$ is the matrix of those variables which are available on a smaller level of geography than what is being used as our aggregate destination. Also, $\mathbf{S}_{DJ} = \left[AREA_J \right]$, $\mathbf{H}_I = \left[1 \ BR_I \ MBR_I \right]$, and

$\mathbf{T}_l = [CRIME_l \quad EDUCEXP_l \quad RESPTAX_l]$. As was discussed in section 4.5.1, our choice of housing attributes is limited by the requirement that they be cross-classified with rent. Census variables that were considered but rejected from the model specification for this and other reasons are number of rooms, year the structure was built, and inclusion of utilities with rent. The bedrooms variable should capture similar influences as number of rooms. In a metropolitan area such as Boston, very little of the housing stock is supplied by new construction, so in order to gauge the quality of housing, it would be more useful to know when a structure was last *remodeled* or *rehabbed* (a question not asked by the Census, probably because of the difficulty in defining what constitutes a sufficient degree of rehabilitation), rather than when the structure was originally *built*. Some preliminary examination of the rental market revealed inclusion of utilities not to have great explanatory power. The influence of this variable may be small, relative to the total monthly rent, and may also suffer from difficulties in definition (that is, which, and how many utilities would have to paid by the landlord in order to qualify for inclusion in this category). These oversimplifications of the housing attributes are acceptable because the focus of this research project is less the Boston housing market than the accessibility patterns implied by where people choose to live and how much they are willing to pay to live in those locations.

Since the modes, m, are motorized and nonmotorized travel, and the level-of-service variables are completely (with the exception of a bias constant) incorporated into the inclusive values from the sub mode choice models. That is $L_{IJ\,motorized} = [\ 1\ IV_{IJ\,motorized}\ ]$ and $L_{IJ\,nonmotorized} = [\ 0\ IV_{IJ\,nonmotorized}\ ]$. $IV_{IJ\,nonmotorized} = \ln\left(e^{\alpha_0 + \alpha_1 ALDIST_{IJ}} + e^{\alpha_2 ALDIST_{IJ}}\right)$, and $IV_{IJ\,motorized}$ is a constant calculated from the motorized sub mode choice model, which is estimated separately.

5.2.4. Housing purchase model. With the exception of the nonmotorized half of the mode choice model system, the hedonic rent regression we are using is identical to that described in Chapter 4:

$$E\left[Value_I^g\right] = \lambda_1^g \ln \sum_J \exp\left(\begin{array}{l} D_J V^g + \rho^g \ln \sum_{m} \exp\left(L_{IJm}\beta^g\right) \\ + \dfrac{1}{\mu_1}\ln B_J + \dfrac{1}{\mu_1}\ln\left(S_{DJ}\gamma_D\right) \end{array}\right).$$
$$+ H_I\lambda_2^g + T_I\lambda_3^g + \frac{1}{\mu_2}\ln B_I + \frac{1}{\mu_2}\ln\left(S_{RI}\gamma_R\right) + \varepsilon_{10} \tag{5-6}$$

Most of the variables in equation 5-6 are identical to the ones used in the rental model. We use the median housing value in a block group, $MEDHV_I^g$, as an instrument for $E[Value_I^g]$, $B_I = 1 + \frac{1}{2}\mu_2^2\lambda_1 Var\left(\tilde{H}_i\right)\lambda_1$, and $S_{RI} = [NOOU_I\ \ NBG_I\ \ AREA_I]$. Since the basic unit of housing, i, is reflected in the totals presented in census tables, it is rather straightforward to construct instruments for $\underset{i\in I}{Var}\left(\tilde{H}_i\right)$ from each census record.

The correction for heterogeneity in destination attributes, $\frac{1}{\mu_1}\ln B_j$, will be omitted because the geographical areas used for aggregate destinations, which I am calling districts, are not much larger than the CTPS zones by which the destination information is available. (In most cases, districts, zones, and census tracts share the same boundaries. However, in a few places, notably the cities of Boston and Brockton, several zones must be combined to match with a combination of census tracts.) Since most of the information about destination heterogeneity would have been lost when destinations were aggregated to zones, the information available from aggregating from zones to districts is not the most useful instrument for total district heterogeneity.

5.3. Results

5.3.1. Motorized sub mode choice model. Recall that modal trip tables were available on the town or neighborhood level, which allow us to simplify the final hedonic rent regression by estimating this section of the model separately. Equation 5-1 shows how the probability of selecting transit over auto for a given trip is related to level-of-service variables. Table 5-3 displays estimates of these parameters and their standard errors, which were produced using the SAS® procedure LOGISTIC.

Note that model 1 produced a counter-intuitive sign for walking times, that is, the more someone has to walk in order to use transit, relative to the

Table 5-3. Estimated Motorized Mode Choice Model

Number of Observations	8138 town-to-town (or Boston neighborhood) pairs
Explained Variable	Selection of transit over auto
Standard errors are reported in parentheses beneath parameter estimates	

Name	Description	Units	Model 1	Model 2
Intercept (β_1)	Transit bias coefficient	--	-1.7689 (0.00616)	-1.6559 (0.00530)
XTCOST (β_2)	Transit fare (NFARE) plus 1/2 of park-and-ride parking cost (PNRCOST), less highway tolls (HTOLL)	1987 ¢	-0.0083 (0.00031)	-0.00106 (0.00003)
HPARK $(-\beta_3)$	Parking cost at destination	1963 ¢ / hour	0.0184 (0.000037)	0.0181 (0.000037)
HDIST $(-\beta_4)$	Highway distance (HDIST; instrument for auto operating costs)	miles	0.0648 (0.000358)	0.0674 (0.00348)
XTWALK (β_5)	Total transit walking time (NTWALK) less auto terminal walking times (HTT)	minutes	0.0161 (0.00448)	--
TWAIT (β_6)	Transit waiting time	minutes	-0.0861 (0.000423)	-0.0855 (0.000423)
XTIVT (β_7)	Transit in-vehicle-time (NTTIME) less auto in-vehicle-time (NHTIME)	minutes	-0.0117 (0.000263)	-0.0117 (0.000262)
TDACC (β_8)	Transit drive access time (zero for walk access paths)	minutes	-0.1586 (0.00109)	-0.1343 (0.000816)

Regression statistics	Model 1	Model 2
-2 ln £ (intercept only)	4,707,095.5	4,707,095.5
-2 ln £ (full model)	3,259,136.3	3,260,437.8

Table 5-3. Estimated Motorized Mode Choice Model, continued

Implied travel values	Units	Model 1	Model 2
Value of in-vehicle-time (β_7/β_2)	1987 $ / hour	8.46	6.62
Value of drive-access to transit time (β_8/β_2)	1987 $ / hour	114.94	75.85
Auto operating cost (β_4/β_2)	1987 ¢ / mile	78.07	63.58
Wait / in-vehicle time factor (β_6/β_7)	--	7.36	7.31
Drive access / in-vehicle time factor (β_8/β_7)	--	13.59	11.45

corresponding auto trip, the more likely this person is to use transit. This result may occur from the way different modes of arrival are treated. Transit trips arriving by auto have lower walking times in general, and of course non-zero drive access times, but in general, travelers who must drive to transit are less likely to use transit. We did not include a drive-to-station dummy variable in our model specification because we suspected this variable would be incorporated by the drive access time variable; instead, it was also confounded with walk access time. Model 2, which omits walk access time, produced parameter estimates with all the expected signs, and parameter values did not differ appreciably from those in model 1.

Since utility units have no readily identifiable interpretation, transportation planners generally prefer to look at the implicit tradeoffs between travel components. This information is presented at the end of Table 5-3 for the reader's convenience.

Conventional wisdom has that passengers value in-vehicle-time at roughly one-half their wage rate for work trips and about one-quarter their wage rate for non-work trips. Since model 2 has $6.62 as the value of an hour of in-vehicle-time, this would imply a wage rate of approximately 13 to 26 dollars in 1987. This wage rate seems reasonable given the proportion of skilled professions and unionized labor in the Boston area.

The implied value of auto operating costs, for which distance driven is an instrument, is about 64 cents per mile. Given that the IRS allows taxpayers to

deduct about 24 cents per mile for business travel in personal autos, one might think that 64 cents per mile is excessive. However, the higher insurance and fuel costs in Massachusetts, relative to the rest of the nation, in combination with the lower expected fuel economy attainable on congested Boston roads, could be partly responsible for the inflated automobile operation costs.

Finally, model 2 predicts that wait time will be perceived as roughly seven times more onerous than in-vehicle-time, and drive access time roughly eleven times worse than in-vehicle-time. Most rules-of-thumb state that these ratios will be in the neighborhood of two to four. CTPS uses a ratio of 2.0 for drive access time when path-building. Also, CTPS adds a 6 to 10 minute wait penalty to each boarding when building transit paths, which on average amounts to a 1.3 to 1.5 penalty ratio (since average waiting times are 20 minutes, as presented in Table 5-2). The aggressive, fast-paced lifestyle of Bay Staters (and most of the East Coast, for that matter) is probably the most important explanation for out-of-vehicle-time being valued as so much more onerous than in-vehicle-time.

5.3.2. Rental model. Resource constraints prevented the estimation of the full rental model. The rental model as described above would require use of the SAS® MIXED procedure, which is available only on mainframes, and not on the DEC workstations at MIT.

5.3.3. Housing purchase model. The housing purchase model was estimated via the nonlinear least squares method, using the SAS® procedure NLIN. This SAS® procedure allows constraints to be placed on the values of parameters, and we generally constrained the signs on destination utilities (the v's) and the travel disutilities (the β's). Constraints were also placed on the size and heterogeneity effect corrections, because discrete choice theory tells us the natural range of these values. The SAS® program used to control the estimation procedure is presented in appendix C as a convenience to future researchers.

Results from the estimation of accessibility from the owner-occupied housing market are presented in Table 5-4. Initial values supplied for parameter estimates were based on earlier model runs described in section 5.4 below, and on intermediate model runs which do not appear in this document.

Note that this model run produced estimates which were suspiciously close to the initial values supplied to the program, and the estimates also had incredibly large standard errors. Such results may suggest scrapping the model entirely. But important insights can be gained by asking why the model performed so poorly. One reason may be that accessibility simply isn't important to homeowners. But the motivation for this thesis is the belief that accessibility is important. Another potential reason is that the employment densities used to model destination utilities are poor instruments for what residents actually find attractive in destinations. Certainly travelers don't have a complete census of regional employment as intuitive knowledge, but

Table 5-4. Estimated Housing Purchase Model

Number of Observations	317 census tracts
Explained Variable	Value of owner-occupied housing
Asymptotic standard errors are reported in parentheses beneath parameter estimates	

Name	Description	Starting Value	Estimates
Accessibility variables			
Access (λ_1)	Hedonic value of accessibility	500	499.998 (153,839)
BOOTSD (v_b)	Destination utility of agricultural, construction, fishing, forestry and mining employment density	0.05	0.0535 (44,477)
FIRED (v_f)	Destination utility of financial, insurance and real estate employment density	0.07	0.0722 (14,192)
GOVTD (v_g)	Destination utility of government employment density	0.03	0.0245 (14,733)
MFGD (v_m)	Destination utility of manufacturing employment density	0.02	0.0233 (6,244)
RETD (v_r)	Destination utility of retail employment density	0.15	0.1486 (7,599)
SVCD (v_s)	Destination utility of service employment density	0.012	0.0124 (2,566)
UTILD (v_u)	Destination utility of communication, transportation, and utilities employment density	0.009	0.009,00 (0.000,1)
WHOLD (v_w)	Destination utility of wholsale employment density	0.007	0.0000 (34,611)
Bias (β_0)	Bias towards motorized travel	0.5	0.501 (2,972)
Inclusive value (β_1)	All modes of motorized travel disutility (IVMOT) or nonmotorized travel disutility (NMDISU)	1	1.000 (183.5)

Table 5-4. Estimated Housing Purchase Model, continued

Name	Description	Starting Value	Estimates
Bias (α_0)	Bias towards walking	-1	-1.008 (19,929)
ALDIST (α_1)	Walking distance	-5	-4.943 (127,735)
ALDIST (α_2)	Bicycling distance	-1	-0.999 (1,682)
Correction term (μ_1)	Scale parameter for destination size effect ($\ln(AREA_j)$)	1	1.000 (507.9)
Other housing variables			
BR (λ_{22})	Average number of bedrooms per house in tract	4,000	4,000 (24,732)
MBR (λ_{23})	Proportion of units in tract with 5 or more bedrooms	12,000	12,000 (128,192)
CRIME (λ_{31})	Violent and property crimes per 1,000 residents	-1,000	-1,000. (354.0)
EDUCEXP (λ_{32})	Educational expenditures per pupil	50	50.00 (11.77)
RESPTAX (λ_{33})	Residential property tax per $1,000 assessed value	500	500. (2,976)
Correction Term (μ_2)	Scale parameter for origin size and heterogeneity effect	1	0.9997 (2,682)
AREA (γ_1)	Area of census tract in 1/1,000 km^2	0.1	0.140 (526,237)
NOOU (γ_2)	Number of owner-occupied units in tract	1	1.395 (5,271,220)

Regression statistics	
Regression degrees of freedom	23
Regression sum of squares	13.927×10^{12}
Residual sum of squares	2.361×10^{12}

somehow travelers are able to process easily obtainable information about the existence and variety of establishments into some notion of the size and attractiveness of an area.

Another reason for the large standard errors may be more related to the functional form. Logsum is a rather smooth function which may also be "flat." It is hard to estimate parameters when shallow gradients result in small step sizes between iterations. The asymptotic correlation matrix supports this inference. Some of the strongest correlations between parameter estimates are between v_b and v_w, and among β_0 and the α's. The negative correlation between v_b and v_w may be due to the similarity in how wholesale enterprises and other enterprises requiring physical labor are perceived, but these types of industry are arbitrarily divided into two categories. Much of the correlation within the α's can be explained by the use of ALDIST for both walking and biking modes.

These strong correlations only frustrate the already difficult task of estimating the accessibility parameters without information about the relative attractiveness among destinations. Recall that our logsum definition of accessibility corresponds to the expected value of the utility of the most attractive destination, net of the disutility to travel there. When the estimation program does not have this attractiveness information, but is free to generate that information from the estimated parameters, we should not be surprised at the loss of statistical efficiency. To overcome this difficulty, we need either destination choice data for travel by all modes, including nonmotorized ones;

or else a different accessibility specification, such as simple summation. We recommend collecting additional data, because of the theoretical reasons against a different specification that were discussed in chapter 4.

The most significant parameter estimates in the full estimated model were those on the town attributes of crime and educational expenditure. This result might suggest that people who purchase homes already have sufficient income which allows them sufficient mobility to maintain their standards of accessibility. Instead, potential home owners are more concerned with issues more closely related to the geography of a plot of land, which aren't as influenced by mobility: is the house in a good school district? is the neighborhood safe?

5.4. Sensitivity to Inputs and Assumptions

One important hypothesis to test is how large a contribution accessibility makes towards housing values. While this questions could be examined by comparing the value placed on typical levels of accessibility with the average value of housing, an equally valid approach is to reestimate the hedonic regression while omitting the accessibility variable. Results of such constrained regressions are presented in Table 5-5.

In each of the three limited regressions presented in Table 5-5, the coefficients on crime rate, educational expenditures, and five or more bedrooms all had

Table 5-5. Housing Purchase Model without Accessibility

Number of Observations	531 census tracts
Explained Variable	Value of owner-occupied housing
Asymptotic standard errors are reported in parentheses beneath parameter estimates	

Name	Description	Model 1	Model 2	Model 3
Intercept (λ_{21})	Constant term	-10,000 (0) (Note 1)	--	-160,143 (29,392)
BR (λ_{22})	Average number of bedrooms per house in tract	-5,992.5 (11,914.2)	-27,587 (13,470)	--
MBR (λ_{23})	Proportion of units in tract with 5 or more bedrooms	189,511 (59,027)	254,467 (65,719)	165,539 (41,994)
CRIME (λ_{31})	Violent and property crimes per 1,000 residents	-748.15 (127.06)	-1,190.5 (140.0)	-827.03 (110.50)
EDUCEXP (λ_{32})	Educational expenditures per pupil	70.33 (5.57)	62.16 (6.89)	69.56 (5.52)
RESPTAX (λ_{33})	Residential property tax per $1,000 assessed value	8,446.2 (1,892.3)	6,598.9 (2,091.7)	9,499.7 (1,885.5)
Correction Term (μ_2)	Scale parameter for size and heterogeneity effect	0.000,1 (0.000,016)	0.707,1 (550.9) (Note 2)	0.000,3 (0.000,031) (Note 3)
AREA (γ_1)	Area of census tract in 1/1,000 km^2	0.000,0 (0.000,000)	0.000,0 (0.000,000)	-- (Note 3)
NOOU (γ_2)	Number of owner-occupied units in tract	0.000,05 (0.000,005)	0.000,0 (0.000,000)	-- (Note 3)
Includes B_1 correction for heterogeneity effect?		No	Yes	No

Table 5-5. Housing Purchase Model without Accessibility, continued

Regression statistics	Model 1	Model 2	Model 3
Regression degrees of freedom	8	8	6
Regression sum of squares	18.289×10^{12}	17.514×10^{12}	18.212×10^{12}
Residual sum of squares	3.115×10^{12}	3.889×10^{12}	3.191×10^{12}

Notes: 1 This regression reported a singular Jacobian matrix, even though the convergence criterion was met. Reported standard errors should be viewed with some skepticism. Likewise, the estimate of the constant term, which did not change from its initial guess, should be viewed with proper caution.

2 The constraint of $\mu_2 \geq \sqrt{2}/2$ was binding.

3 In this regression, AREA was used as the only origin size variable, and therefore it is not possible to separately identify μ_2 and γ_1.

the expected sign and reasonable magnitudes. It is difficult to develop intuition for the sign on the coefficient for residential property tax; property tax may be viewed as a burden on home owners, or as a proxy for local services. Since the sign on property tax is positive, we can assume that its role as a proxy for local services dominates. Another factor may be that high property taxes are correlated with center cities, which are more attractive for reasons not explained by the reduced model, accessibility included.

The sign of the coefficient on the number of bedrooms is even more curious. One explanation for the negative sign on bedrooms (and in the constant term) might be the great attention placed on other variables such as educational expenditures. Note that sixty (a conservative approximation of λ_{32}) times the roughly \$3,000 per year spend for education on average is essentially equal to the \$180,000 average home price in the Boston area. Another confounding factor may be studio condominiums in well-to-do neighborhoods such as the Back Bay, which would have a high value, but would also lower the average number of bedrooms per unit in an area.

Time and other resource constraints prevented a more thorough examination of all assumptions and sensitivities. Questions that should be addressed in future research include whether accessibility parameters and hedonic prices vary between geographic sub markets within Eastern Massachusetts, how parameters vary among socioeconomic groups, and the functional form of the hedonic housing price model.

Chapter 6

Policy Implications and Conclusions

During the course of this project, the normal course of exploration continued to reveal the richness of study relating to accessibility. Each day would uncover some new subtlety which would need to be addressed in some manner or assumed away. Many times it was frustrating not to rise to the mental challenge presented by some issue, but the reality is that project sponsors need results, and even researchers must pay attention to other parts of their lives. Most of the suggestions that we have brainstormed for future improvements to the accessibility methodology presented in this thesis fall into one of two categories: improvements to better understanding accessibility itself, and incorporating better accessibility measures into other transportation models.

6.1. Improving the Understanding of Accessibility

The most obvious way of increasing general knowledge of accessibility is to estimate accessibility models on different sources of data, in order to compare and contrast results. For example, not only can accessibility be imputed from housing rents, but it can also be calculated from models of destination choice, or even estimated from special surveys which may present residents with hypothetical travel and destination situations. These different data collection designs may also help us answer questions about

how long-run and short-run perceptions of accessibility differ, if at all. Different types of variables, such as the locations of groceries and pharmacies we wanted to include, but weren't able, will help us better understand what factors actually determine the attractiveness of destinations.

Another important question is have people's accessibility responses (the parameters estimated in chapter 5) remained relatively constant over time, or are they subject to fluctuation due to changes in income and taste, and innovations of new products and services. A researcher may want to examine historical travel and development patterns to answer these questions. A similar issue is cross-city comparisons: can accessibility measures developed for Boston be used to evaluate the quality of life in Los Angeles, for example? Or will such an exercise merely a futile discovery of self-selection bias — one would find that Bostonians much prefer living in Boston to L.A., which is why they live in Boston to begin with!

At one point we were also interested in the kinds of accessibility firms would be interested in (before we discovered how difficult a task it was describing the accessibility that residents want). Extending the accessibility model to firms would be quite easy. Firms derive "utility" from profit-making potential. It may even be easier to do the accounting in dollars and avoid the fiction of "utility units." Firms would then be expected to trade off access to customers with that of access to material inputs and labor markets.

The accessibility framework would also benefit from behavioral scientists who can better describe the data gathering process. Certainly residents don't purchase entire copies of census files, yet they manage to find those destinations important to meeting their everyday needs. Such research will help us find more realistic data sources, as well as address questions about how technology innovations will affect accessibility. Improvements in communication and computer technology occur rather frequently, and may replace the need to actually travel to a location in order to gather information about it.

6.2. Accessibility in Other Transportation Models

Figure 4-1 is a representation of all the decisions affected by accessibility which transportation planners attempt to model. Accessibility's use is far more broad than merely a project evaluations tool, but accessibility is also important to understanding location decisions, travel demand, and the impact of travel demand on the performance and capacity of the transportation system. In turn, the outcome of influences on these systems affects accessibility in the future.

When using accessibility as an evaluation tool, it is tempting to hypothesize what land use changes may result from a change in accessibility. While skilled judgment and intuition are valuable assets, a more empirical treatment of land use responses to accessibility can only strengthen one's case. This thesis has paid more attention to relations between accessibility

and existing housing. Just as important are new housing starts and desires for rezoning which occur from changes in accessibility. Examination of these location markets should also give insight to whether a simultaneous-equations bias exists in the current accessibility model.

Since accessibility is one means of studying the interactions between transportation and land use, it makes sense that in addition to being interested in location decisions that we are also interested in travel demand. On a short run basis, travel demand is affected by accessibility to activities and by specific needs. However, in the long run, we don't know specific consumption needs, but can only speak of travel frequencies for a specific purpose, or of an "average travel day." Another important distinction with respect to accessibility is that actual travel is constrained by time and money budgets. (These constraints were relaxed to calculate an accessibility index.) However, insight might be gained from considering the analog between accessibility and actual travel:

Accessibility (Potential Travel) = Destination Utility + Travel Disutility

Actual Travel = Destination Choice + Mode & Route Choice

Due to the linear nature of both the utility specification and the resource constraints, the utility optimization problem for determining actual travel may not have a unique solution. (This is also caused by not having enough constraints.) One way to resolve this problem is to treat actual travel as

having a probability distribution itself. For example, if we assume actual travel is proportional to potential travel (accessibility), we get the result predicted by the exponential gravity model. The exponential gravity model can be shown to be the expected value of this actual travel distribution. However, we could just as easily assume that travel only occurs to the closest (or farthest) destinations. Therefore, it would be most useful to know the variance of the actual travel distribution.

Any discussion of travel demand will eventually consider transportation capacity constraints. Likewise, our examination of accessibility has assumed virtually unlimited capacity, which we know to be unfortunately untrue. For example, in-vehicle time of autos and buses is affected by the number of vehicles using those streets. This is typically modeled by the BPR equation:

$$LinkTime = FreeFlowTime\left(1 + \alpha\left(\frac{LinkVolume}{LinkCapacity}\right)^{\beta}\right)$$

Also, all transit vehicles suffer a degradation in travel time as greater numbers of passengers board and alight. Assuming acceleration is unaffected by passenger weight for constant stop spacing, one might write another model as

Transit IVT = β_0 + β_1 *Hwy Link Time* + β_2 *Boardings and Alightings*

Furthermore, as the number of passengers in a transit vehicle increases, comfort decreases as passengers have to stand, and possibly be near an undesirable person.

However, congestion and load factors are often highly dependent on the time of day. Since a person making a long run location decision would not consider such detail as the time of day he or she might make any possible type of trip, an average level of congestion might be appropriate.

Obviously a large number of research opportunities exists in even this narrow subset of transportation research. Almost certainly there are other issues regarding accessibility that the author has neglected to mention which would also be beneficial projects.

References Consulted

Alonso, William (1964) *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press, Cambridge, Mass.

Anas, Alex (1985) "The Combined Equilibrium of Travel Networks and Residential Location Markets." *Regional Science and Urban Economics*. Vol. 15, p. 1-21.

Aschauer, David Alan (1989) "Can Underfunded Highways and Transit Stall the Economy?" Women's Transportation Seminar, Greater Chicago Area Chapter, pamphlet, based on the keynote address at the forum "Moving America ... Developing a National Transportation Policy," also sponsored by the WTS chapter, on October 17.

Association for Public Transportation, Inc. (1992) *Car-Free® in Boston*. Eighth edition.

Barrett, Frank A. (1974) *Residential Search Behavior: A Study of Intra-Urban Relocation in Toronto*. Department of Geography, Atkinson College, York University, Ontario, Canada.

Ben-Akiva, Moshe and Steven R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Mass.

Birch, David L., Reilly Atkinson, III, Richard P. Coleman, William L. Parsons, Kenneth Rosen, and Arthur P. Solomon (1974) *Models of Neighborhood Evolution*. Joint Center for Urban Studies, Cambridge, Mass.

Brand, Daniel (1973) "Theory and Method in Land Use and Travel Forecasting." *Highway Research Record 422*, p. 10-20. Highway Research Board, National Research Council, Washington, D.C.

Brand, Daniel (1990) "Research Needs for Analyzing the Impacts of Transportation Options on Urban Form and the Environment." *Transportation, Urban Form, and the Environment: Proceedings of a Conference*. Special Report 231, Transportation Research Board, National Research Council, Washington, D.C.

Brand, Daniel, Brian Barber, and Michael Jacobs (1967) "Technique for Relating Transportation Improvements and Urban Development

Patterns." *Highway Research Record 207*. Highway Research Board, National Research Council, Washington, D.C.

Bureau of the Census. (1987) 1987 Economic Censuses CD-ROM Volume 1: Report Series, Release 1D (machine-readable data files). Washington, D.C.

Bureau of the Census. (1992) Census of Population and Housing, 1990: Summary Tape File 3 on CD-ROM (machine-readable data files). Washington, D.C.

Burtt, Everett J., Jr. (1967) *Plant Relocation and the Core City Worker*. United States Department of Housing and Urban Development.

Butler, E. W., F. Stuart Chapin, Jr., G. C. Hemmens, E. J. Kaiser, M. A. Stegmen, and S. F. Weiss (1969) "Moving Behavior and Residential Choice: A National Survey." National Cooperative Highway Research Program Report 81, Highway Research Board, Washington, D.C.

Butler, Richard Vernon (1977) *Hedonic Indexes of Urban Housing: Theory and Problems of Cross-Sectional Estimation*. Ph. D. Thesis, MIT.

Chapin, F. Stuart, Jr. (1974) *Human Activity Patterns in the City: Things People Do in Time and Space*. John Wiley and Sons, New York.

Clark, Mildred, ed. (1973) *Behavioral Demand Modeling and Valuation of Travel Time*. Transportation Research Board Special Report 149. National Research Council, Washington, D.C.

DeGroot, Morris H. (1989) *Probability and Statistics*, 2nd ed. Addison-Wesley Publishing Company, Reading, Mass.

Dewees, D. N. (1978) *Changes in Transportation Performance and Accessibility in Metropolitan Toronto: 1964-71*. Research Paper No. 100, Centre for Urban and Community Studies, University of Toronto. Diandas, J. (1984) "Alternative Approaches to Transport in Third World Cities: Issues in Equity and Accessibility." *Ekistics*, Vol. 51, p. 197-212.

Dickey, John W. (1983) *Metropolitan Transportation Planning*, 2nd ed. Hemisphere Publishing Corporation, New York.

Domencich, Thomas A., Gerald Kraft, and Jean-Paul Valette (1968) "Estimation of Urban Passenger Travel Behavior: An Economic Demand Model." *Highway Research Record 238*, p. 64-78. Highway Research Board, National Research Council, Washington, D.C.

Domencich, Thomas A. and Daniel McFadden (1975) *Urban Travel Demand: A Behavioral Analysis*. North-Holland Publishing Company, Amsterdam.

Ermuth, Frederick (1974) *Residential Satisfaction and Urban Environmental Preferences*. York University, Downsview, Ontario.

Ettema, Dick, Aloys Borgers, and Harry Timmermans (1993) "Toward a Production System Model of Activity Scheduling Behavior." TRB Preprint.

Fredland, Daniel R. (1974) *Residential Mobility and Home Purchase: A Longitudinal Perspective on the Family Life Cycle and the Housing Market*. D.C. Heath and Company, Lexington, Mass.

Fukushige, Mototsugu and Michio Hatanaka (1991) "Estimation of a Regression Model on Two or More Sets of Differently Grouped Data." *Journal of Econometrics*, Vol. 41, p. 207-226.

Garden, Jackie and Erik Mackinlay (1980) "Report of a Round Table Meeting held in Den Haag, 16-17 January 1980." Service Levels of Urban Public Transport: a series of documents on a study commissioned by the Ministry of Transport and Public Works of the Netherlands, II.

Greene, William H. (1990) *Econometric Analysis*. MacMillan Publishing Company, New York.

Groves, Randall Dean (1990) *A Study of the Importance of Intangible Attributes in the Marketing of Computer Workstations*. M.S. Thesis, MIT.

Haitovsky, Yoel (1973) *Regression Estimation from Grouped Observations*. Griffin's Statistical Monographs and Courses, No. 33. Hafner Press, New York.

Ham, John and Cheng Hsiao (1984) "Two-Stage Estimation of Structural Labor Supply Parameters Using Interval Data from the 1971 Canadian Census." *Journal of Econometrics*, Vol. 24, p. 133-158.

Hamer, Andrew Marshall (1976) *The Selling of Rapid Rail Transit: A Critical Look at Urban Transportation Planning*. Lexington Books, Lexington, Mass.

Hammer, Philip G., Jr. and F. Stuart Chapin, Jr. (1972) "Human Time Allocation: A Case Study of Washington, D.C." Technical Monograph, Center for Urban and Regional Studies, University of North Carolina at Chapel Hill.

Hansen, W. G. (1959) "How Accessibility Shapes Land-Use." *Journal of the American Institute of Planners*. Vol. 25, p. 73-76.

Hanson, S. and M. Schwab (1987) "Accessibility and Intraurban Travel." *Environment and Planning A*, Vol. 19, p. 735-748.

Hensher, David A. and Peter R. Stopher, eds. (1979) *Behavioral Travel Modelling*. Redwood Burn Limited, Trowbridge and Esher, Great Britain.

Hornor, Edith R., editor (1989) *Massachusetts Municipal Profiles*. Information Publications, Palo Alto, Calif.

Hunt, David T., Adriene B. Bailey, Christina Adidjaja, and J. Douglas Carroll (1984) "Using the 1980 Census to Evaluate the Equity of Transit Service in Northern New Jersey." *Transportation Research Record 992*, p. 67-72. Transportation Research Board, National Research Council, Washington, D.C.

Ingram, D. R. (1971) "The Concept of Accessibility: A Search for an Operational Form." *Regional Studies*, Vol. 5, p. 101-107.

Institute of Public Administration and Netherland's Economic Institute (1980[a]) "First Interim Report." Service Levels of Urban Public Transport: a series of documents on a study commissioned by the Ministry of Transport and Public Works of the Netherlands, III.

Institute of Public Administration and Netherland's Economic Institute (1980[b]) "Annexes to the First Interim Report." Service Levels of Urban Public Transport, a series of documents on a study commissioned by the Ministry of Transport and Public Works of the Netherlands, III A.

Johnson, Norman L. and Samuel Kotz (1970) *Distributions in Statistics: Continuous Univariate Distributions-1.* Houghton Mifflin Company, Boston. Johnson, Norman L. and Samuel Kotz (1972) *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley and Sons, Inc., New York.

Kanafani, Adib (1983) *Transportation Demand Analysis.* McGraw-Hill Book Company, New York.

Karlqvist, Anders, Lars Lundqvist, Folke Snickars, Jörgen W. Weibull, eds. (1978) *Spatial Interaction Theory and Planning Models.* (*Studies in Regional Science and Urban Economics*, Vol. 3. Åke Andersson and Walter Isard, eds.) North Holland Publishing Company, Amsterdam.

King, A. Thomas and Peter Mieszkowski (1973) "Racial Discrimination, Segregation, and the Price of Housing." *Journal of Political Economy*, Vol. 81, No. 3 (May/June), p. 590-606.

Kunert, Uwe (1993) "Weekly Mobility of Life Cycle Groups." TRB Preprint.

La Gory, Mark and John Pipkin (1981) *Urban Social Space.* Wadsworth Publishing Company, Belmont, Calif.

Lerman, Steven R. (1975) *A Disaggregate Behavioral Model of Urban Mobility Decisions.* Ph.D. Thesis, MIT.

Lim, Gill C. (1980) "Growth Control and Price Variations in Urban Housing Markets in a Developing Country." Princeton University. (prepared for the 24th International Geographical Congress, Tokyo)

Linneker, B. J. and N. A. Spence (1992) "Accessibility Measures compared in an analysis of the Impact of the M25 London Orbital Motorway on Britain." *Environment and Planning A*, Vol. 24, p. 1137-1154.

Maddala, G. S. (1977) *Econometrics.* McGraw-Hill Book Company, New York.

Manski, Charles F. (1977) "The Structure of Random Utility Models." *Theory and Decision*, Vol. 8, No. 3 (July), p. 229-254.

Massy, William F., Ronald E. Frank, and Thomas Lodahl (1968) *Purchasing Behavior and Personal Attributes.* University of Pennsylvania Press, Philadelphia.

Mitchell, Robert B. and Chester Rapkin (1954) *Urban Traffic: A Function of Land Use*. Columbia University Press, New York.

Moore, Eric G. (1972) *Residential Mobility in the City*. Association of American Geographers, Washington, D.C.

Moorthy, K. Sridhar (1991) "Measuring Overall Judgements and Attribute Evaluations: Overall First vs. Attributes First." Technical working paper, Marketing Science Institute, Cambridge, Mass.

Morris, J. M., P. L. Dumble, and M. R. Wigan (1979) "Accessibility Indicators for Transport Planning." *Transportation Research A*, Vol. 13, p 91-109.

Niedercorn, J. H. and B. V. Bechdolt, Jr. (1969) "An Economic Derivation of the 'Gravity Law' of Spatial Interaction." *Journal of Regional Science*. Vol. 9, No. 2, p. 273-282.

Rossi, Peter H. (1980) *Why Families Move*, 2nd ed. Sage Publications, Beverly Hills.

SAS® Institute, Inc. (1985) *SAS® User's Guide: Basics, Version 5 Edition*. Cary, N.C.

SAS® Institute, Inc. (1989) *SAS/STAT® User's Guide, Version 6, Fourth Edition*, Vol. 1 and 2. Cary, N.C.

Schnare, Ann B. and Raymond J. Struyk (1976) "Segmentation in Urban Housing Markets." *Journal of Urban Economics*, Vol. 3, No. 2 (April), p. 146-166.

Simmons, James W. (1968) "Changing Residence in the City - A Review of Intra-Urban Mobility." *Geographical Review*, Vol. 58, p. 622-651.

Strandberg, Keith W. (1992) "Fare Collection: Systems Now and Future." *Mass Transit*, Vol. 19, p. 47-52.

Straszheim, Mahlon R. (1975) *An Econometric Analysis of the Urban Housing Market*. National Bureau of Economic Research, New York.

Strathman, James G., Kenneth J. Dueker, and Judy S. Davis (1993) "Effects of Travel Conditions and Household Structure on Trip Chaining." Center

for Urban Studies, School of Urban and Public Affairs, Portland State University. TRB Preprint.

Ting, Harold Montford (1971) "Aggregation of Attributes for Multiattributed Utility Assessment." Technical Report No. 66, Operations Research Center, MIT.

Wallace, J. P. and A. Sherret (1973) "Estimation of Product Attributes and Their Importance." *Lecture Notes in Economics and Mathematical Systems, Econometrics* Vol. 89.

Weisbrod, Glen E. (1978) *Determinants of Residential Location Demand: Implications for Transportation Policy.* S.M. Thesis, MIT.

Weisbrod, Glen E., Robert J. Parcells, and Clifford Kern (1984) "A Disaggregate Model for Predicting Shopping Area Market Attraction." *Journal of Retailing,* Vol. 60, No. 1 (Spring), p. 65-83.

Wheaton, William (1993) MIT course 11.433J/15.021J (Real Estate Economics) lecture notes.

Whyte, William Hollingsworth (1988) *City: Rediscovering the Center.* Dell Publishing Group, Inc., New York.

Wingo, Lowdon, Jr. (1961) *Transportation and Urban Land.* The Johns Hopkins Press, Baltimore, Md.

Wolpert, J. (1965) "Behavioral Aspects of the Decision to Migrate." *Papers and Proceedings of the Regional Science Association.* Vol. 15, p. 159-172.

Appendix A

Calculating Intra-District Nonmotorized Distances

Intrazonal trips and skims present a difficulty to most transportation planners because of the assumption of all activity within a zone occurring at a centroid. One such difficulty is that using the geographic coordinates of centroids for calculating walking and biking distances would result in zero distance for intrazonal trips. Clearly neighborhoods don't exist as single, dense points, so we would expect some finite distance for trips within a zone or district.

One potential source of an instrument for reasonable intra-district distances is the Census, which has data on land area. Larger districts would have larger expected distances between any two activities. However, this expected distance would also be influenced by the shape of the district; thin, snaky districts would have longer intra-district travel distances than districts which more resemble squares or circles. We choose to examine a square district in order to simplify calculations, and as a compromise shape — circles are more compact, but most real districts will be less compact, so considering them as a square will underestimate the intra-district distance.

Figure A-1 shows such a prototypical square district, with axis labels. Assuming uniform development — which is admittedly unrealistic, but the

best assumption we can make with the information available — then the

expected   distance   of   an   intra-district   trip,   q,

is $E[q] = \frac{1}{d^4} \int_0^d \int_0^d \int_0^d \int_0^d \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \, dx_1 dx_2 dy_1 dy_2$,   which   cannot   be

evaluated analytically. (Note that x and y are the coordinates of the

endpoints of a hypothetical trip, and that $d = \sqrt{Area}$ or the length of each side of the square district.) We can reduce the complexity of this problem if we consider the expected value of horizontal and vertical displacements, E[s], separately; calculate q* as the representative distance from the expected displacements; and then examine Jensen's inequality to see which way using q* will bias our estimates

Figure A-1. Distances within a Prototypical District

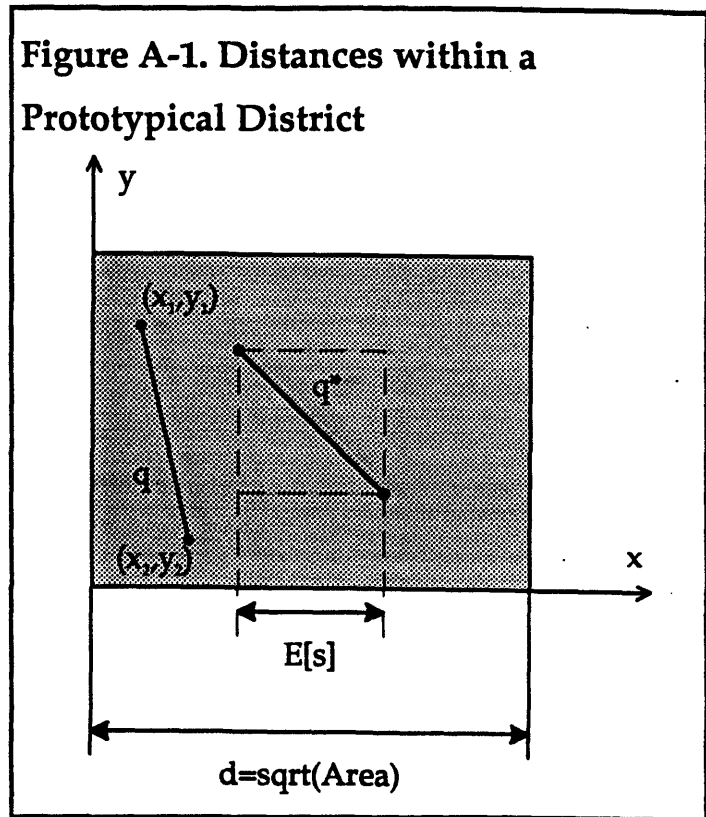of E[q]. Since the distance formula, $\sqrt{\Delta x^2 + \Delta y^2}$, is concave up, q* will underestimate E[q]. This is acceptable because we expect the use of airline distances to underestimate true travel distances as well. By Pythagorus, $q* = \sqrt{2}E[s]$. By considering a uniform activity density along a line segment of length d, Use as legs of a right triangle, we can show that the expected

value between any two activities, E[s] is $\int_0^d \int_0^d \frac{1}{d^2}|x_1 - x_2| dx_1 dx_2$. By symmetry,

$$E[s] = 2\int_0^d \int_{x_1}^d \frac{x_2 - x_1}{d^2} dx_2 dx_1 = \frac{2}{d^2}\int_0^d \left(\frac{d^2}{2} - x_1 d + \frac{x_1^2}{2}\right) dx_1 = \frac{2}{d^2}\left(\frac{d^3}{6}\right) = \frac{d}{3} \text{ so } q* = \frac{\sqrt{2Area}}{3}.$$

Appendix B

Model Extension for Transit Mode-of-Arrival

Transportation analysts often model mode-of-arrival to transit stations, especially for purposes of comparing alternatives for major capital investment in a single transit line. This appendix describes how the accessibility model can be expanded to utilize mode-of-arrival data.

Transit users have a choice of whether to walk, bike, or drive to their first station or stop. Variables which affect the mode-of-arrival decision are travel time to the station (which will be correlated for walk and bike travel, since we assume they can use the same sidewalks, just as we did in the nonmotorized travel sub mode choice model), parking costs, and differences in fare which might arise if a Park-and-Ride lot isn't available at the closest station.

The "reach transit" model is then

$$V_{\text{Walk to Transit}} = \gamma_1 NTWTT,$$

$$V_{\text{Bike to Transit}} = \gamma_2 + \gamma_3 NTBTT + \gamma_6 \left(NFARE_{\text{Bike to Transit}} - NFARE_{\text{Walk to Transit}} \right), \text{ and}$$

$$V_{\text{Drive to Transit}} = \gamma_4 + \gamma_5 TDACC$$
$$+ \gamma_6 \left(PNRCOST + NFARE_{\text{Drive to Transit}} - NFARE_{\text{Walk to Transit}} \right),$$

where V's represent the systematic utilities of each method of reaching transit. NFARE, TDACC, and PNRCOST have the same meaning as in chapter 5. NTWTT and NTBTT refer to network-generated walk- and bike-to-transit times. The inclusive value of reaching transit is

$$IV_{transit} = \ln\left(\begin{array}{c} e^{\gamma_1 NTWTT} + e^{\gamma_2 + \gamma_3 NTBTT + \gamma_6 (NTFARE_{Bike} - NTFARE_{Walk})} \\ + e^{\gamma_4 + \gamma_5 NTDACC + \gamma_6 (NTPARK + NTFARE_{Drive} - NTFARE_{Walk})} \end{array}\right),$$

which would appear as an extra variable describing the systematic utility of a transit trip.

Appendix C

SAS® Code to Estimate Accessibility Parameters

A SAS® program similar to the following was used to estimate the accessibility model in this document. (Minor differences include making the variable names appearing here consistent with those in chapter 5, and cleaning up some of the comments.)

```
options  pagesize=60 linesize=80;
libname  mustang '/cts/sramming';

/* ****************************************
this file is Models/nloov.ssp

T 12 Oct 93
Scott Ramming

estimate the Non Linear Owner Occupied housing
market regression with accessibility,
and Variance as an instrument for the
heterogeneity effect
**************************************** */

/* ----------------------------------------------------------------
The SAS data set mustang.oo4v has the following format:

origin destination odflag ij-pair variables origin-specific variables

   1       1          1      <list>           <list>
   1       2          0      <list>           <repeated list>
   1       3          0      <list>           <repeated list>

  ...

   1      776         0      <list>           <repeated list>
   1      777         9      <list>           <repeated list>
   2       1          1      <list>           <new list for origin 2>
   2       2          0      <list>           <repeat of list for
                                                          origin2>

  ....
```

```
The variable odflag was created in the data step using
if-then statements which evaluate the special operators
``first.'' and ``last.'' for by-variables

odflag = 1 for the first observation of an origin
         9 for the last observation of an origin
         5 for the first, last, and only observation of an origin
             (a case we hope doesn't occur)
         0 otherwise
----------------------------------------------------------------- */


proc nlin data=mustang.oo4v maxiter=25;

/* Initial parameter guesses */
parameters ll=500 vb=0.05 vf=0.07 vg=0.03 vm=0.02
                  vr=0.15 vs=0.12 vu=0.09 vw=0.07
                  p=1 b0=0.5 b1=1 b2=1 a0=-1 a1=-5 a2=-1 m1=1
                  l21=-8000 l22=4000 l23=12000
                  l31=-1000 l32=100  l33=20
                  m2=1 gr1=0.1 gr2=0.01 gr3=1;


/* Set constraints on parameters based on discrete choice theory
   and economic intuition */

bounds m1>1e-20, m2>0.7071, gr1>1e-20, gr2>1e-20, gr3>1e-20,
       vb>1e-20, vf>1e-20, vg>1e-20, vm<1e-20, vr>1e-20,
       vs>1e-20, vu>1e-20, vw>1e-20, b2>1e-20, lam>1e-20;


/* Now do some tricky stuff so we can
   logsum acc over all destinations   */

if (odflag=1)or(odflag=5) then do;
    /* Set accumulator variables to zero */
    sacc=0; /* sums IJ access */
    sab=0;  /* IJ access weighted boots for d medhv / d v boots */
    saf=0;  /*                         fire                     */
    sag=0;  /*                         govt                     */
    sam=0;  /*                         mfg                      */
    sar=0;  /*                         ret                      */
    sas=0;  /*                         svc                      */
    sau=0;  /*                         util                     */
    saw=0;  /*                         whol                     */

    saut=0; /* IJ access weighted travel disutility for d medhv/ d p */
    saum=0;  /* weird IJ access weighted func of motorized util */
             /* for d medhv / d b0 */
    samum=0; /*                         b1 */
    sanm=0;  /*                         b2 */
    sanmw=0; /*                         a0 */
    sanmdw=0; /*                        a1 */
    sanmdb=0; /*                        a2 */
```

```
        sala=0;  /* IJ access weighted - (1/ m1**2) ln area */
                /* for d medhv / d m1                      */
end;

/* now set some temporary variables */

    euwalk=exp( a0 + a1*aldist );
    eubike=exp( a2* aldist );
    enmdisu=  euwalk + eubike;
    eunm=exp( b1 * log (enmdisu) );
    eum= exp( b0 + b1 * ivmot );    /* ivmot = IV motorized */
    eivdisu= eum + eunm;
    eijacc= exp( bootsd*vb + fired*vf + govtd*vg + mfgd*vm  +
                 retd*vr   + svcd*vs  + utild*vu + whold*vw +
                 p*log(eivdisu) + (1/m1) * log(ddarea) );


                                    /* ddarea = AREA J */


/* now accumulate variables over all destinations */

    sacc + (eijacc);
    sab + (bootsd*eijacc);
    saf + (fired*eijacc);
    sag + (govtd*eijacc);
    sam + (mfgd*eijacc);
    sar + (retd*eijacc);
    sas + (svcd*eijacc);
    sau + (utild*eijacc);
    saw + (whold*eijacc);

    saut   + (eijacc*log(eivdisu));
    saum   + (eijacc*((p*eum)/eivdisu));
    samum  + (eijacc*((p*eum*ivmot)/eivdisu));
    sanm   + (eijacc*((p*log(enmdisu)*eunm)/eivdisu));
    sanmw  + (eijacc*((p*eunm*b2*euwalk)/(eivdisu*enmdisu)));
    sanmdw + (eijacc*((p*eunm*b2*aldist*euwalk)/(eivdisu*enmdisu)));
    sanmdb + (eijacc*((p*eunm*b2*aldist*eubike)/(eivdisu*enmdisu)));
    sala   + (eijacc*(-1/m1**2)*log(ddarea));

/* a few more auxillary variables */

    lvl= 122**2 * vbr + 2*122*123*covbr + 123**2 * vmbr;
    /* lvl = lambda2' Var(H I) lambda2 */
    bi= 1 + 0.5* m2**2 * lvl;

    model medhv = 11*log(sacc) +
                  121          + br*122        + mbr*123 +
                  crime*131 + educexp*132 + resptax*133 +
                  (1/m2)*log(bi) +
                  (1/m2)*log(area*gr1 + nbg*gr2 + own*gr3);
```

```
/* now one last trick */
if (odflag=0) or (odflag=1) then do;
    _resid_=0;_loss_=0;_weight_=0;_wgtjpj_=0; end;
else do;
    _resid_=(medhv-model.medhv);_loss_=_resid_**2;
    _weight_=1;_wgtjpj_=1; end;


/* This will get SAS to ignore all observations except for the last
   one of each origin, which has the logsum totals of ijAccess stored
   in its accumulator variables                                      */

/* now specify derivatives for hill climber */

der.ll=log(sacc);
der.vb=ll*sab/sacc;
der.vf=ll*saf/sacc;
der.vg=ll*sag/sacc;
der.vm=ll*sam/sacc;
der.vr=ll*sar/sacc;
der.vs=ll*sas/sacc;
der.vu=ll*sau/sacc;
der.vw=ll*saw/sacc;

der.p=ll*saut/sacc;
der.b0=ll*saum/sacc;
der.b1=ll*samum/sacc;
der.b2=ll*sanm/sacc;
der.a0=ll*sanmw/sacc;
der.a1=ll*sanmdw/sacc;
der.a2=ll*sanmdb/sacc;
der.m1=ll*sala/sacc;

der.l21=1.;
der.l22=br  + (m2 *(l22*vbr+l23*covbr))/bi;
der.l23=mbr + (m2 *(l22*covbr+l23*vmbr))/bi;
der.l31=crime;
der.l32=educexp;
der.l33=resptax;

der.m2= m2*lvl/bi - (1/m2**2)*log(bi)
        -(1/m2**2)* log(area*gr1 + nbg*gr2 + own*gr3);
der.gr1= (1/m2) * area/(area*gr1 + nbg*gr2 + own*gr3);
der.gr2= (1/m2) * nbg/ (area*gr1 + nbg*gr2 + own*gr3);
der.gr3= (1/m2) * own/ (area*gr1 + nbg*gr2 + own*gr3);

run;
```

Colophon

Text and headers are printed in 12 pt Palatino. Sub- and superscripts are 8 pt Palatino. In text, subscripts are 3 pt below the baseline and superscripts are 5 pt above the baseline. Chapter headings are printed in 18 pt Palatino Bold. Section headings, and titles of tables and figures are 14 pt Palatino Bold. Figures are surrounded by a 2 pt line frame. Kibo recommended that I use Aldus since that was the typeface Herman Zapf designed as the companion text face for the display face Palatino; however, only Palatino was available on the Laser Writer IINT used to produce this document.

This document was produced using Word for Windows Version 2.0a. It does not use "straight" or "square" quotation marks, nor does it use the word "paradigm" (although "framework" is probably just as bad). Unfortunately, I was unable to use fi, ffi, fl, ffl, etc. ligatures, because of the inherent short-sightedness of the PostScript system. The original version ("running set") of this document was printed by an Apple Laser Writer IINT. This document was formatted with wider inner margins to facilitate double-sided copying and binding.

Kibo also told me that colophons are expected to have a grape leaf at the end, so here it is: ⚜