A DISAGGREGATE BEHAVIORAL MODEL
OF URBAN MOBILITY DECISIONS

by

STEVEN RICHARD LERMAN

B.S.C.E., Massachusetts Institute of Technology
(1972)
S.M., Massachusetts Institute of Technology
(1973)

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
at the
Massachusetts Institute of Technology
June, 1975

Signature of Author ..........................................................
Department of Civil Engineering, May 2, 1975

Certified By ..........................................................
Thesis Co-Supervisor

..........................................................
Thesis Co-Supervisor

Accepted By ..........................................................
Chairman, Department Committee on Graduate
Students of the Department of Civil Engineering

MASS. INST. TECH.
JUN 20 1975
ABSTRACT

A BEHAVIORAL MODEL OF URBAN MOBILITY DECISIONS

by

STEVEN RICHARD LERMAN

Submitted to the Department of Civil Engineering on May 2, 1975 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The close relationship between transportation policy and changes in urban form have long been recognized by transportation planners. This study develops a model of one aspect of that relationship, the household's choice of residential location, housing, auto ownership and mode to work. These decisions, termed the mobility choices, are viewed as being completely interdependent and as determined jointly by each household rather than independently or in a sequence.

In this study, existing theories of residential location are explored and extended to provide a more realistic representation of the mobility decisions. The theoretical development provides a consistent framework for the exploration of a range of hypotheses about the way in which various factors determine the mobility choices.

A number of multinomial logit models are estimated to reflect these hypotheses using data from the 1968 Washington, D.C. Home Interview Survey. Separate models are developed for single and multiple worker households in order to represent the differences in the mobility choice process between the two groups.

The model is used to explore in a preliminary way the potential impacts of a variety of alternative transportation policies designed to reduce automobile usage. However, the most significant result of this study lies in the possibility of using the basic modelling approach to reformulate existing comprehensive land use models to better reflect the causal mechanisms which determine the spatial distribution of activities in cities and the impacts of public policy on that distribution. A revised forecasting process in which auto ownership and work trip travel patterns are predicted in the land use forecasting step is recommended.

Thesis Co-Supervisors: Wayne M. Pecknold Moshe Ben-Akiva

Titles: Associate Professor of Civil Engineering Assistant Professor of Civil Engineering
ACKNOWLEDGEMENTS

A great number of people have provided assistance to me in the course of this study. Professors Moshe Ben-Akiva and Wayne Pecknold served as thesis co-supervisors, and each gave valuable advice on various aspects of the work. In addition, Professors Marvin Manheim and Gregory Ingram were members of my doctoral committee and provided many useful ideas and suggestions.

A number of fellow students also provided advice during the course of this research. In particular, I would like to thank Frank Koppelman, James Hudson, Thomas Adler, Uzi Landau, and Len Sherman, all of whom were sources of information and ideas at various times.

The data for this study was originally collected by the Metropolitan Washington Council of Governments. In addition, a processed data file prepared by Richard Albright of Cambridge Systematics proved to be an important resource.

Drafts of this work were typed by Charlotte Goldberg, Patricia Otis, Rebecca Lord and Donna Martin. The final version was typed by Deborah S. Sylvain and Sonja Arsove.

The actual research was supported by a fellowship from the Motor Vehicle Manufacturers Association. Their generous financial contribution is gratefully acknowledged. In addition, early stages of my work were funded by the Sloan Foundation.

Finally, I wish to acknowledge the intangible contribution of my wife Lori, to whom this thesis is dedicated.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION AND SUMMARY</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.1 Transportation, Urban Location and Related Mobility Decisions</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.2 Overview of the Study</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>1.3 Summary of Major Findings and Recommendations</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.4 Summary of Remaining Chapters</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>THE BEHAVIORAL STRUCTURE OF URBAN MOBILITY AND TRAVEL CHOICES</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>2.1 Introduction</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>2.2 Mobility Choice Models and Policy Analysis</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.3 The Behavioral Unit</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>2.4 Mobility and Travel Choices: A Hierarchical Framework</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2.5 Modelling Implications of the Choice Hierarchy</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>2.6 Dimensions of the Mobility Bundle</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>2.7 Factors Affecting Mobility Decisions</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>2.8 Summary of Chapter 2</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>LITERATURE REVIEW</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3.1 Structure of the Literature Review</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3.2 Theoretical Studies of Urban Mobility</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>3.3 Sociological Studies of Urban Location</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>3.4 Impact Oriented Studies</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>3.5 Comprehensive Urban Land Use Models</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>a-econometric models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b-Lowry-based models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c-the NBER Urban Simulation Model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d-the Lee critique</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.6 Behavioral Models of Mobility Decisions</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>a-unidimensional models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b-sequential models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c-joint models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d-summary of previous models</td>
<td></td>
</tr>
<tr>
<td>3.7 Summary and Conclusion of Literature Review</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 4: A BEHAVIORAL THEORY OF MOBILITY CHOICE</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>4.1 A Behavioral Analysis of Mobility Decisions</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>4.2 An Extension of the Basic Alonso Model: Residential Bid Rent Curves</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>4.3 Non-CBD Workplaces and Spatial Opportunities in the Alonso Model</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>4.4 Travel Time and Congestion</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>4.5 Urban Services: The Tiebout Hypothesis</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>4.6 Pollution</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>4.7 The Role of Race and Ethnicity</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>4.8 Households with Multiple Workers</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>4.9 Households with no Workers</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>4.10 Households without Fixed Workplaces</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>4.11 Bid Rent Curves and the Full Mobility Choice</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>4.12 Summary of Chapter 4</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 5: ISSUES IN CHOICE THEORY MODELS OF MOBILITY</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>5.2 Models of Choice Behavior</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>5.3 The Logit Model</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>5.4 Previous Applications of Choice Theory to Predicting Transportation Demand</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>5.5 Maximum Likelihood vs. Maximum Score Estimation</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>5.6 Recursive vs. Joint Estimation of Choice Models</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>5.7 The Problem of Grouped Alternatives</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td>5.8 Requirements for Consistency of Grouped Alternatives</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>5.9 A Choice Model with Grouped Alternatives</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>5.10 Assumptions of the Grouped Model</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>5.11 Physical vs. Attribute Grouping</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>5.12 Grouping Mobility Alternatives</td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>5.13 Summary of Chapter 5</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 6: DESCRIPTION OF THE DATA</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>6.2 Perceived vs. Measured Attributes</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>6.3 Data Sources</td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>6.4 Definition of Subsample</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>6.5 Profile of the Washington, D.C. Data Set</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>6.6 Summary of Chapter 6</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 7: MODEL SPECIFICATION AND ESTIMATION RESULTS</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>7.1 Specification of Behavioral Models</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>7.2 Measures in the Joint Model Specification</td>
<td>213</td>
<td></td>
</tr>
<tr>
<td>7.3 Specification of the Joint Utility Function</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.4</td>
<td>The Set of Available Alternatives</td>
<td>243</td>
</tr>
<tr>
<td>7.5</td>
<td>The Conditional Location/Housing Model for Single Worker Households</td>
<td>248</td>
</tr>
<tr>
<td>7.6</td>
<td>The Joint Mobility Choice Model for Single Worker Households</td>
<td>252</td>
</tr>
<tr>
<td>7.7</td>
<td>The Conditional Location/Housing Model for Multi-Worker Households</td>
<td>256</td>
</tr>
<tr>
<td>7.8</td>
<td>The Joint Mobility Choice Model for Multi-Worker Households</td>
<td>260</td>
</tr>
<tr>
<td>7.9</td>
<td>Summary of Estimation Results</td>
<td>266</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER 8: BEHAVIORAL IMPLICATIONS OF THE MODELS</strong></td>
<td>269</td>
</tr>
<tr>
<td>8.1</td>
<td>Evaluating the Implications of Behavioral Models</td>
<td>269</td>
</tr>
<tr>
<td>8.2</td>
<td>Comparison of Coefficients</td>
<td>272</td>
</tr>
<tr>
<td>8.3</td>
<td>Pure Alternative Effects</td>
<td>278</td>
</tr>
<tr>
<td>8.4</td>
<td>Value of Time</td>
<td>283</td>
</tr>
<tr>
<td>8.5</td>
<td>Prototypical Household Analysis: The Base Case</td>
<td>288</td>
</tr>
<tr>
<td>8.6</td>
<td>Prototypical Household Analysis: Policy Scenarios</td>
<td>301</td>
</tr>
<tr>
<td>8.7</td>
<td>Summary of Behavioral Implications</td>
<td>314</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER 9: CONCLUSIONS AND RECOMMENDATIONS</strong></td>
<td>317</td>
</tr>
<tr>
<td>9.1</td>
<td>Summary of the Thesis</td>
<td>317</td>
</tr>
<tr>
<td>9.2</td>
<td>Implications of the Research for a Revised Analysis Framework</td>
<td>319</td>
</tr>
<tr>
<td>9.3</td>
<td>Recommendations for Future Research</td>
<td>325</td>
</tr>
<tr>
<td>9.4</td>
<td>Conclusion</td>
<td>328</td>
</tr>
<tr>
<td></td>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>329</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.1</td>
<td>Impact Oriented Econometric Models</td>
<td>61</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of Previous Mobility Models</td>
<td>84</td>
</tr>
<tr>
<td>6.1</td>
<td>Distribution of Primary Worker's Occupation</td>
<td>195</td>
</tr>
<tr>
<td>6.2</td>
<td>Distribution of Household Income by Occupation</td>
<td>197</td>
</tr>
<tr>
<td>6.3</td>
<td>Number of Workers vs. Household Size With Row Percentages and Means</td>
<td>199</td>
</tr>
<tr>
<td>6.4</td>
<td>Housing Type Choice in Cluster Sample (With Sample Percentages)</td>
<td>201</td>
</tr>
<tr>
<td>6.5</td>
<td>Automobile Ownership vs. Mode of Primary Worker for Owners (With Sample Percentages)</td>
<td>203</td>
</tr>
<tr>
<td>6.6</td>
<td>Automobile Ownership vs. Mode of Primary Worker for Renters (With Sample Percentages)</td>
<td>204</td>
</tr>
<tr>
<td>6.7</td>
<td>Automobile Ownership vs. Mode of Primary Worker for Households in Single Family Dwellings (With Sample Percentages)</td>
<td>205</td>
</tr>
<tr>
<td>6.8</td>
<td>Automobile Ownership vs. Mode of Primary Worker for Households in Multi-Family Dwellings (With Sample Percentages)</td>
<td>206</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary of Measures Used in the Model</td>
<td>225</td>
</tr>
<tr>
<td>7.2</td>
<td>Definition of Variables</td>
<td>240</td>
</tr>
<tr>
<td>7.3</td>
<td>Variables and Choices to Which They Pertain</td>
<td>243</td>
</tr>
<tr>
<td>7.4</td>
<td>Estimates for Conditional Location/Housing Model: Single Worker Households</td>
<td>249</td>
</tr>
<tr>
<td>7.5</td>
<td>Estimates for Joint Mobility Model: Single Worker Households</td>
<td>253</td>
</tr>
<tr>
<td>7.6</td>
<td>Estimates for Conditional Location/Housing Model: Multi-Worker Households</td>
<td>261</td>
</tr>
<tr>
<td>7.7</td>
<td>Estimates for Joint Mobility Model: Multi-Worker Households</td>
<td>264</td>
</tr>
<tr>
<td>8.1</td>
<td>Coefficient Estimates for Parameters With Variation Across Household Types</td>
<td>275</td>
</tr>
<tr>
<td>8.2</td>
<td>Values of Time ($/hr.)</td>
<td>289</td>
</tr>
<tr>
<td>8.3</td>
<td>Attributes of Hypothetical City</td>
<td>293</td>
</tr>
<tr>
<td>8.4</td>
<td>Level of Service in Hypothetical City</td>
<td>295</td>
</tr>
<tr>
<td>8.5</td>
<td>Base Case Prototypical Households</td>
<td>296</td>
</tr>
<tr>
<td>8.6</td>
<td>Impacts for Base Case</td>
<td>299</td>
</tr>
<tr>
<td>8.7</td>
<td>Impacts for Policy 1</td>
<td>305</td>
</tr>
<tr>
<td>8.8</td>
<td>Impacts for Policy 2</td>
<td>306</td>
</tr>
<tr>
<td>8.9</td>
<td>Impacts for Policy 3</td>
<td>307</td>
</tr>
<tr>
<td>8.10</td>
<td>Impacts for Policy 4</td>
<td>308</td>
</tr>
<tr>
<td>8.11</td>
<td>Impacts for Policy 5</td>
<td>309</td>
</tr>
<tr>
<td>8.12</td>
<td>Impacts for Policy 6</td>
<td>310</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>The Proposed Forecasting Process</td>
<td>18</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Hierarchy of Choice</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Three Stage Hierarchy of Choice</td>
<td>37</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>The Lowry Model Family Tree</td>
<td>69</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>The Mobility Sequence in the NBER Urban Simulation Model</td>
<td>71</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Consumer Equilibrium for Land and Composite Good</td>
<td>93</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Bid Rent Curves</td>
<td>98</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>A Feasible Rent Surface for Two Groups</td>
<td>99</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Bid Rent Curves for Non-CBD Workplace</td>
<td>102</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Bid Rent Curves for a Regional Activity Center</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Congested Travel Speed Near CBD</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Bid Rent Curves for Uniform and Congested Travel Times</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Bid Rent Curves for Different Jurisdictions</td>
<td>112</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Bid Rent Curves With a Pollution Source</td>
<td>115</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Bid Rent Curves for Positive Discrimination</td>
<td>120</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Bid Rent Curves for Negative Discrimination</td>
<td>122</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Multiple Worker Households With a Primary Worker</td>
<td>125</td>
</tr>
<tr>
<td>Figure 4.13</td>
<td>Bid Rent Curves for Two Worker Household; Both Workers with Fixed Workplace</td>
<td>127</td>
</tr>
<tr>
<td>Figure 4.14</td>
<td>Bid Rents When Workplace is Indeterminate</td>
<td>130</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Error Due to Truncation of Taylor Expansion</td>
<td>167</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Probability Error Due to Truncation</td>
<td>169</td>
</tr>
<tr>
<td>Figure 8.1</td>
<td>Pure Auto Ownership Utilities</td>
<td>280</td>
</tr>
<tr>
<td>Figure 8.2</td>
<td>Pure Alternative Housing Probabilities</td>
<td>282</td>
</tr>
<tr>
<td>Figure 8.3</td>
<td>Zone System for Hypothetical City</td>
<td>291</td>
</tr>
<tr>
<td>Figure 9.1</td>
<td>The Traditional Forecasting Process</td>
<td>320</td>
</tr>
<tr>
<td>Figure 9.2</td>
<td>A Revised Forecasting Process</td>
<td>323</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION AND SUMMARY

1.1 Transportation, Urban Location, and Related Mobility Decisions

The city, defined as an entire urban area rather than an arbitrary jurisdiction, is a complex system which can be viewed from a number of perspectives. To the economist, the city is a region where economies of scale result from an agglomeration of activities; to the sociologist, the city is a sophisticated social system where individuals and groups interact on a level impossible in agrarian societies; to the urban planner, the city is all too frequently an area where the lack of insight into underlying causal mechanisms leads to misdirected policy. From each of these perspectives, the transportation system plays a critical role in determining the urban form and the interactions which take place. In the simplest terms, goods and services are moved between points of production and residences, and people are moved between residences and workplaces. At a more sophisticated level of abstraction, goods and services are distributed through what is in actuality a complex web of activities, and people travel to a great number of destinations to fulfill a variety of physical and psychological needs. In U.S. cities, these transportation activities can absorb more than half of all urban land, and frequently produce some of the most critical problems of public policy.
In the field of transportation planning, the methodological tools used to address major policy problems have for the most part been large scale, comprehensive urban land use, or urban activity system, models. These models have been developed at great expense, yet most often fail to produce either behavioral insights into the determinants of location patterns or useful forecasts (Lee, 1973). Furthermore, as urban problems grow more and more complex and the policies proposed to alleviate them grow more sophisticated, these models are frequently unable to provide any policy sensitive forecast. In most cases, the failure of existing comprehensive urban activity system models as useful policy analysis tools can be attributed to their lack of a consistent and empirically verified behavioral theory about how each decision-making unit determines its location and related choices. The models tend to rely on correlations between exogenous variables and location rather than a realistic representation of behavior. More causal models have generally been theoretical in nature, and are not useful in forecasting contexts.

For this reason, a new approach towards modelling urban form based on more causal reasoning seems to be needed. This study is a first step in this direction in that it develops a set of models which represent one major component of urban location patterns, the household's location, housing, automobile ownership, and mode to work decision, in a behaviorally consistent way. This group of choices, termed the mobility decision, are viewed as jointly, or simultaneously determined by the household.
The objective of this study is not to construct a complete urban land use model. Obviously, mobility choice models alone would be insufficient, since other actors, most notably firms, landlords and developers, make decisions which affect urban location. Rather, the study has the following objectives:

(1) To develop a consistent and credible theory of how households make their mobility choices and how those choices relate to their daily travel decisions;

(2) To structure and estimate a joint, or simultaneous, disaggregate choice model of household mobility decisions which reflects the theory developed;

(3) To use the models developed to examine a number of hypotheses about mobility decisions and to explore in a preliminary way the impact of alternative policies;

(4) To provide a basis for restructuring the current transportation planning process to reflect a more behavioral view of how urban location and travel patterns are determined.

1.2 Overview of the Study

In order to meet the above objectives, it is necessary to construct a general framework for considering how household mobility decisions and the day-to-day travel choices are related. This framework, termed the choice hierarchy, distinguishes between two sets of choices: the long run
mobility decisions of location, housing, auto ownership and mode to work and the short run travel choices of frequency, mode, destination, route and time of day for non-work trips. These decisions are assumed to be structured hierarchically, i.e. location and related decisions are made with travel choices indeterminate and travel choices are then made conditional on the outcome of the mobility choice process. Within each group of choices, decisions are assumed to be made by a joint process in which the full range of possible tradeoffs is considered by the household.

The joint nature of realted groups of decisions has been recognized since the earliest attempts to model the choice process at the level of the decision-making unit. In a study of mode choice for work trips, Warner (1962) explicitly states that,*

"the decision regarding choice of mode is interdependent with the choice of location of home and the choice of car ownership!"

Warner conveniently sidesteps this issue by stipulating that his analysis deals only with mode choice decisions conditional on the other aspects of the mobility decision.

More recently, the perspective of viewing the entire spectrum of mobility and travel choices as part of a general choice hierarchy has formed the basis for the development of joint behavioral choice models for non-work trip purposes (Ben-Akiva, 1973; Adler and Ben-Akiva, 1975) and has been partially implemented by Lerman and Ben-Akiva (1975) in the estimation of a joint choice model of automobile ownership and mode to work. However, this study represents the first attempt to fully employ

*abstracted from Kain (1964)
the choice hierarchy perspective in representing the long run decisions associated with the full set of urban mobility decisions.

The choice hierarchy has the basic advantage of avoiding assumptions of an arbitrary sequence of decisions which underlie existing comprehensive urban land use models. However, it poses a significant modelling problem since the number of mobility alternatives open to every household is very large and each option is characterized by a great number of attributes, not all of which are known or measurable. The analytic approach used in this study to resolve these problems is the application of random utility models, one of a more general class of choice theory models (Manski, 1975).

Random utility models rely on a clear, household level theory of behavior and explicitly account for the observer's imperfect information about the decision process being represented. Furthermore, because these models use information about the basic decision-making unit, they avoid the aggregation of data and the possible fallacies which such aggregation can cause (de Neufville and Stafford, 1971). The particular analytic form selected, the multinomial logit model, was chosen for a variety of practical and theoretical reasons, including the lack of alternative methods for modelling decision problems with large choice sets and the substantial base of successful prior applications which exists.

In this study, the theoretical development of the mobility decision process begins with the basic concepts of location theory of urban
economics. In a qualitative way, a broad range of behavioral hypotheses about mobility choices is explored. These hypotheses give rise to a set of possible causal factors which should be represented in the models. Using data from Washington, D.C., a number of choice models were estimated to reflect these hypotheses.

Different models were developed for single and multiple worker households in order to represent the essential behavioral differences between the two groups which were hypothesized to exist. These differences are most significant in the household's perception of the work trip attributes of alternative residential locations. For single worker families, only one relevant work trip destination exists, while for multi-worker households some combination of all the work trips must be considered.

The models estimated should be viewed as preliminary in nature, since budget limitations imposed unavoidable restrictions in sample size and the scale of the study. Nevertheless, even within relatively small samples, important causal variables which frequently have been ignored in other models emerge as significant determinants of the household mobility choice process.

The final models estimated are used in a number of scenarios to explore the possible impacts of a variety of policies on a set of "prototypical" households. While these tests provide no conclusive evidence
about the total aggregate response to various changes in the transportation system, they do provide some important insights into the possible magnitude of the aggregate response and how various socioeconomic groups might be differentially impacted by alternative policies.

1.3 Summary of Major Findings and Recommendations

The basic findings and recommendations arising from the study are as follows:

(1) It is feasible to model the entire mobility decision consisting of location, housing, automobile ownership and mode to work as a joint choice process. The models developed using this methodology are consistent with a priori hypotheses about the nature of the household mobility choice process and avoid the need for arbitrary assumptions about a sequence of household decisions.

(2) Even with very small sample sizes, the use of a joint choice model produces statistically efficient parameter estimates due to the explicit treatment of the full range of tradeoffs a household can make. More conventional conditional decision models in which some aspects of the choice process are assumed to be fixed do not make complete use of the available information in the estimation process.

(3) The mobility choice process for households with more than one full-time worker appears to be far more complex than has been assumed in other studies. Model specifications for single worker households which appear reasonable both theoretically and empirically are unacceptable for multi-worker households. Assumptions about the existence of a primary
worker, or "breadwinner", appear to be questionable, since household location and other mobility choices may well depend on the work trip pattern of all the workers in the household. The results of this study do not offer any simple resolution to this problem, but do suggest that a major source of error in our understanding of mobility choice may well arise from an inadequate conceptual and empirical representation of multi-worker households. Since the proportion of such households has grown markedly in recent years and will in all probability continue to do so, it seems that an improved understanding must ultimately be developed before substantial progress in land use modelling can be made.

(4) Most of the broad range of causal factors which have been postulated to affect mobility choice behavior appear to be verified as significant determinants of mobility choice. This range of variables includes the usual location theory measures of the attributes of the work trip, prices, and land but extends to descriptors of local taxes, urban services, neighborhood quality, racial composition, density, and accessibility to shopping by both car and transit. Furthermore, the behavior of households may vary considerably depending on their socioeconomic condition as described by their income, household size, race, number of licensed drivers, and family structure.

(5) It is possible to view each potential location alternative as a distinct housing unit without actually estimating a model at that highly disaggregated level. It can be theoretically demonstrated that under the assumption that the multinomial logit model applies to the highly disaggregated alternatives, it is possible to group similar alternatives
(i.e. alternatives with nearly identical measurable attributes) and still obtain consistent estimates of the household's evaluation of the disaggregated alternatives. This proof allows for a behavioral mobility choice model to be estimated from existing data rather than requiring detailed information about each true alternative. Furthermore, the resulting model of grouped alternatives is actually a version of the multinomial logit form and can easily be estimated.

(6) The current approach used in forecasting urban travel patterns should be revised to be consistent with the proposed choice hierarchy. Existing practice generally does not recognize that work travel patterns arise directly from urban land use decisions and should be treated as such. The proposed process, depicted in Figure 1.1, relies on joint disaggregate choice models of household behavior. Non-work travel is forecasted after the pattern of location has been determined, while the work trip pattern is an output of what has traditionally been termed urban land use.

(7) The use of this proposed process opens a number of possibilities for resolving the problems associated with making aggregate forecasts from disaggregate models. This so-called "aggregation problem" is the object of fairly intensive research by McFadden and Reid (1975), Westin (1973), and Koppelman (1975). One critical issue in resolving this problem is the need to develop the distribution of the independent
Figure 1.1

THE PROPOSED FORECASTING PROCESS
variables in the choice model at the zonal level. By developing a consistent and behaviorally structured chain of disaggregate choice models, distributional information from one forecasting step can be used to make consistent aggregate forecasts in the next.

(8) There are systematic and explainable differences between the coefficient estimates for single and multiple worker households. In general, because single worker households often have an adult member at home throughout the day, they are more concerned with locational attributes than their multi-worker counterparts. Also, the desired level of auto ownership of single worker households is much more sensitive to the number of licensed drivers they have. This is probably because multi-worker households can share the use of an auto for work trips while single worker families tend to have mutually exclusive needs for cars.

(9) A preliminary analysis of alternative transportation policies designed to reduce car usage indicates that while many policies have significant impacts on locational preferences, incremental approaches maintain the basic urban spatial patterns currently observable. Major large scale changes in demand which are qualitatively different from the base case are achieved only through relatively large changes in the level of service and extent of the public transit system.

This analysis, however, is quite limited in that only the potential shifts in demand patterns are measured. More conclusive policy implications must ultimately be derived from a more complete model system which explicitly accounts for both short and long run supply responses.
1.4 Summary of Remaining Chapters

The remainder of this thesis can be conceptually divided into two sections, the theoretical development and the model results, comprising Chapters 2 through 5 and Chapters 6 through 8 respectively.

In Chapter 2, a comprehensive framework for considering household mobility and travel choices is proposed. A variety of important theoretical issues including the motivation for the structuring of behavioral models, the relevant decision-making unit for mobility choices, the relevant dimensions of the mobility decision, and a typology of causal factors which might enter into a mobility choice model, are considered.

Chapter 3 is a review of some of the existing literature which is relevant to this study. In this review, a broad range of perspectives on mobility decisions is examined. Since it would be hopeless to compile an exhaustive review, an attempt is made to synthesize the most relevant aspects of prior studies.

In the following chapter, a more rigorous development of alternative causal hypotheses is attempted. This analysis is an extension of the urban location theories of Alonso (1964) and others and extends their basic logic to consider a more realistic range of behavioral factors which affect mobility choice. The basic logical construct used to formulate these hypotheses is the bid rent curve, the surface of the greatest amount a household would be willing to pay for land at each point in the city.
Chapter 5 is a discussion of the choice theory methods used in the empirical study. The basic elements of choice theory and the multinomial logit model are derived in this chapter. However, the greatest emphasis is placed on the resolution of a number of particular methodological problems, including alternative estimation procedures, the use of sequential or joint model structures, the use of attribute representation of mobility alternatives rather than the physical units themselves and the grouping of alternatives.

This completes the theoretical portion of the dissertation. Chapter 6 then describes the data used in the empirical study, with particular emphasis on explicitly stating the assumptions used to develop the complete data base. In addition, a number of cross-tabulations are presented in order to explore some of the attributes of the data set in a preliminary way.

Chapter 7 describes the model structure and the estimation results for both single and multiple worker households. A discussion of the implications of these results follows in Chapter 8, in which the implied values of time are derived and the potential demand impacts of some simple but illuminating policy scenarios are examined.

Chapter 9 is a summary of the work and a discussion of the basic conclusions and recommendations of the study.
Chapter 2

THE BEHAVIORAL STRUCTURE OF URBAN MOBILITY AND TRAVEL CHOICES*

2.1 Introduction

This chapter presents a comprehensive framework for describing the interrelationships among various aspects of a household's transportation-related decision-making. This framework includes not only the urban mobility choices of location, housing, automobile ownership and mode to work which are the principal focus of this thesis, but also the conventional travel decisions such as trip frequency, destination, and mode for non-work travel with which transportation planners have traditionally been concerned. Thus, the general framework spans the entire range of traveller behavior, from the long term choice of location to the day to day trip-making for shopping, personal business, social and recreational purposes.

Because this study is primarily concerned with forecasting the quantitative effects of transportation policy on the various aspects of mobility choice, the theoretical development will focus on economic models of consumer behavior rather than its social or psychological aspects. This orientation is not means to imply that questions such as the sociological and psychological effect of mobility decisions are not at all important; rather, it is based on the practical hypothesis that such "non-economic" considerations can be meaningfully incorporated into

*Portions of this chapter have appeared in an earlier form in a report entitled A Behavioral Model of Automobile Ownership and Mode of Travel by Cambridge Systematics, Inc. (1974) authored by Ben-Akiva and Lerman.
the way in which those making mobility decisions evaluate the costs and benefits of the alternatives they face. Thus, the role of neighborhoods and other sociological factors are treated in this development as aspects of the mobility decisions which are evaluated as costs and benefits in the most general sense.

The remainder of this chapter will consider the following questions:

(1) Why should behavioral models of mobility and travel choice be developed?

(2) What is the relevant decision-making unit for mobility choices?

(3) What are the important components of the mobility choice, and how do they relate to one another?

(4) How are the conventional transportation decisions of trip frequency, destination, mode, time of day and route related to the mobility decisions?

(5) What are the modelling implications of the overall behavioral theory?

(6) What mobility options are relevant for any one decision-maker?

(7) What are the causal factors which determine the mobility choice?

2.2 Mobility Choice Models and Policy Analysis

This thesis is primarily concerned with the development and estimation of models of location, housing, automobile ownership and mode to work. This group of decisions, termed in Chapter 1 the mobility choices, is of natural interest to the urban transportation planner since it determines the pattern of urban land use which the transportation system serves. However, before beginning the detailed theoretical and methodo-
logical discussion which is included in the following chapters, it is useful to address the more general issues underlying the development of the behavioral models of mobility choice in this thesis and some of the basic limitations of those models.

Models, be they theoretical, empirical, or qualitative, are a means of expressing hypotheses about a phenomenon in a way which is logically consistent and presumably less complex than the phenomenon itself. Thus by definition a model is an abstraction and simplification of the real world. Once developed, models provide a means of evaluating alternative changes in the current situation which arise from exogenous factors or explicit policies by various actors in the process being modelled.

In transportation planning, the most commonly used models are empirically estimated. Traditionally, the developers of these models have paid only limited attention to capturing the essentials of the behavioral process and have devoted the major portion of their analytic effort to making detailed, though inaccurate, forecasts. This is particularly the case with urban land use models, which have relied on weak causal arguments and non-behavioral model structures (Lee, 1973).

In this thesis, a very different perspective is adopted. The empirical models developed in this study focus almost entirely on representing the behavioral decision process. In order to do so, the entire modelling methodology relies on a choice theory perspective. The analytic details of this methodology will be considered in Chapter 5.
At this point, it suffices to state that choice theory is concerned with the behavior of an individual decision-maker confronted with a set of alternatives from which one and only one must be selected. The decision-maker is assumed to assess each option with a utility function and select the option from which he derives the greatest utility. However, because there is always some uncertainty in the measurement of this utility function, it is impossible for an observer to know which alternative the decision-maker will select; only the probability of selection for each alternative can be determined. By observing the actual choices, or revealed preferences, of the decision-makers, estimates of the parameters of the utility function can be obtained which can then be used for forecasting.

There is implicit in the use of an empirically estimated model to make forecasts the hypothesis that the underlying preferences which the utility function represents are reasonably stable over the time period of interest. This hypothesis of temporally stable behavior is at the heart of any demand analysis. Without it, reliable estimates of a model's parameters can still be obtained, but any attempt to draw inferences about the future impacts of alternative policies are meaningless. In fact, the need for stability of a model is one of the foremost reasons for approaching any analysis from a causal or behavioral perspective. As will be discussed in Chapter 3, the Literature Review,
many previous models have relied on correlations between observed variables and urban mobility decisions rather than on a consideration of underlying behavior. It is far more likely that these correlations will change rather than the underlying behavioral decision mechanism. Thus, a model which is firmly rooted in a behavioral theory will be more likely to provide reliable estimates of how various policies will affect location and related decisions than a more naive, correlative model. This is particularly true for policies which will make radical changes in the transportation system because such changes will in all probability alter the statistical correlations upon which non-causal models are based.

Another basic limitation of the models developed in this thesis is that they represent a static view of mobility decisions. In reality, mobility decisions are made in a dynamic, or time-dependent way. For example, households rarely consider altering their location, housing, auto ownership and mode to work all at once. Instead, various components of the mobility choice may be adjusted at different times. Furthermore, the adjustment process is not without both monetary and time costs. Relocation within an urban area requires substantial effort and typically involves a major search of the housing market. A household which has not relocated for a number of years may not be in the position which maximizes its utility; the cost to the household of readjusting its location to demographic or price changes may be higher than the
benefits which it perceives it would gain. Other aspects of the mobility decisions can be adjusted with somewhat lower cost. Automobiles can be bought or sold with fairly low transaction cost, and the choice of mode to work can be changed from one day to the next.

The nature of the interactions among choices in this dynamic process is extremely complex and not well understood. Varying transaction costs probably result in households making different mobility choices in different time frames. For example, mode to work might be reconsidered by the household whenever any other mobility choice is altered. Similarly, a change in housing choice might lead to a readjustment in both auto ownership and mode to work.

In this study, this complex dynamic process is modelled as a static one, principally due to the lack of adequate time series data needed to estimate any dynamic model. In the static structure, the entire set of mobility choices is considered jointly, since in the long run a household can modify all aspects of its choice, regardless of the time frame. While it is unlikely that every household in a cross-section sample will be observed to select the alternative which maximizes its utility, the comprehensive treatment of the entire mobility decision including auto ownership and mode to work allows households to come fairly close. For example, if a household would actually find a more expensive location with improved access more desirable, it can often own more autos to compensate.
A further limitation a static model structure imposes is that any policy conclusions derived from it are restricted to a discussion of the final equilibrium response rather than the process of adjustment towards equilibrium. For this reason, the models presented in this thesis should be viewed as representations of the long term effect of policies on mobility choices. For most purposes, however, it is unlikely that this limitation will be of major importance. Most transportation policies which are likely to influence location patterns require extremely long implementation periods. Major urban transit and highway systems require from five to ten years between initial planning and final implementation and may last for as long as thirty to fifty years. Thus, even though the period of adjustment for mobility decisions may be quite long, the period over which the major policies being evaluated have effect is probably much greater. Transient effects, while of interest, are likely to be relatively minor compared to the total impact of major transportation policies.

2.3 The Behavioral Unit

The first facet of a theoretical development of mobility behavior that must be considered is the basic behavioral unit which makes mobility decisions. Clearly, aggregate analyses relying on zonal, tract or district mobility choices are inappropriate in a behavioral analysis. In the same way that travel demand modellers have begun to recognize that "zones do not commute - people commute," (McFadden and Reid, 1975) zones do not
make mobility decisions. However, a more subtle issue is whether the behavioral unit for the entire set of mobility decisions is the individual or the household and how a household can be meaningfully defined.

It seems fairly clear that in most cases location and housing decisions are made by households. However, households can be defined in a number of ways. In the U.S. Census (U.S. Department of Commerce, 1973) a household is defined as all individuals sharing a dwelling unit. Alternatively, Lansing and Mueller (1964) examined location and auto ownership decisions of "spending units," defined as all individuals sharing a dwelling unit who pooled most of their income and made joint economic decisions. Neither of these definitions is entirely satisfactory. In the former case, two roommates sharing rent would be considered as a single household, even though they reached their location decisions as a result of two somewhat independent choice processes. In the latter case, two individuals living together who are related but each working (e.g., two sisters living together) might be two spending units; however, it is conceivable that they reached their location decisions through a single decision-making process.

In order to restrict the focus of this study, a household will be defined as one or more individuals who share a dwelling unit and make joint economic decisions. Dwelling units shared by more than one spending unit are explicitly not considered.
Mode to work, and to a lesser extent automobile ownership might reasonably be considered decisions made by individuals rather than by households. In many cases, individuals within a household have complete use of one of the autos in that household. In a sense, that automobile is "owned" by the individual. However, if the household is narrowly defined as one or more individuals sharing a residence and making joint economic decisions, the individual's "ownership" of the automobile impacts directly on the remaining household members. Aside from the obvious possibility of a remaining member using the auto as either a passenger or driver, the household has also allocated a portion of its income to the purchase, maintenance and other costs of owning the car. In the case of the choice of mode of travel, if one worker takes a car to work, the remaining drivers in the household no longer have use of it. Thus, while individuals are making mode choice decisions, the outcomes of those decisions within a household are highly interdependent.

The use of the household as the behavioral unit is therefore an abstraction; for households of more than one person, the household's mobility decision is in reality the result of an extremely complex bargaining process among household members, each with their own preferences and relative bargaining strengths. However, to explicitly incorporate the effect of individual household members on the household's ultimate mobility choice would require a substantially more sophisticated theory of
group interaction than presently exists. Furthermore, this added detail would not be likely to provide significantly more insight into the effects of various policies than would a household level theory.

2.4 Mobility and Travel Choices: A Hierarchical Framework

In order to develop a behavioral theory at the household level, it is useful to consider two general classes of transportation-related decisions. (Ben-Akiva, 1973) The first class, termed mobility choices in Chapter 1, consists essentially of the long run decisions which households make. These choices include employment location, residential location, housing, and automobile ownership. Such decisions are not generally made on a daily or weekly basis; instead, they are made relatively infrequently. In addition, mode to work is included, since this trip is made every workday and is likely to directly be reconsidered simultaneously with any change in the other decisions.

The second class of decisions, termed travel choices, include the choices made by the household with respect to non-work trips. Examples of travel choices include frequency, mode, destination, time of day, and route for shopping, recreational, social, and other non-work travel purposes. These choices are of a more short run nature and can be altered rather frequently. Furthermore, it is hypothesized that the particular non-work trips a household makes are not re-evaluated jointly with the mobility choices.
Within each of these two groups of decisions the choices are highly interdependent. A household's choice of residential location should be viewed as an element of the mobility decision, or in other words, as made jointly with all other mobility choices. In the same manner, the choice of travel mode for a shopping trip, for example, should be viewed as an element of a shopping travel decision.

The travel choices are assumed to be directly conditioned by the mobility choices. The outcome of the mobility choice determines the options available to the household for the travel choices. Thus, mobility and travel decisions are made hierarchically, the latter conditioned on the former.

The mobility choices determine all the travel choices for the work trip except for possible short term adjustments of route and perhaps time of day for employees without fixed working hours. As will be further discussed in Chapter 4, the trip from home to work and return is treated as part of the mobility decision because it is the most important trip made by the household, and generally has a greater influence on other mobility choices than other trip purposes. Indeed, a household's evaluation of alternative locations is not independent of the usual mode to work decision. Thus, although the choice of mode to work may be made on a different time frame than the choice of residential location, it is a higher-level choice with respect to travel decisions for other trip purposes. The choices with respect to trips for non-work purposes, with the
possible exclusion of school trips, have a much shorter response time than the mobility choices, including the work trip.

The assumed hierarchy of the mobility and travel choices is depicted in Figure 2.1. Each block in the figure represents a group of highly interdependent choices which are assumed to be jointly determined by the household. In other words, within the context of a static representation of a dynamic process the household is assumed to consider the full range of possible tradeoffs when contemplating a change in any one of the components. For example, the household may choose to live in the downtown area and own zero or one auto, or it may select a suburban dwelling with two cars. Similarly, in making its travel choices a household may trade-off the costs and benefits of travelling to a small neighborhood store or suburban shopping center by car against those of using transit to the central city.

The outcome of this decision process is assumed to be determined by the household in a two stage procedure. First, the household makes long term decisions by selecting a mobility alternative. Then, conditioned on the selected mobility alternative, the household makes travel choices for a variety of non-work travel purposes. Taking each block as a single choice, the mobility and travel decisions are made by a conditional decision process; however, within each block the choices are made by a joint decision process.
Figure 2.1

HIERARCHY OF CHOICE
Obviously, the proposed choice hierarchy does not apply to every household. Some households may, for example, not make any work trips; for them, the work trip decision is irrelevant. Other households may have no licensed drivers, and they therefore never choose to own autos. For this reason, it is necessary to define market segments, groups of households with the same underlying choice structure and with nearly homogeneous utility functions.*

In the empirical portion of this study, only two distinct market segments, single and multi-worker households, are defined. However, an attempt is made to eliminate households for whom the choice hierarchy may not be a realistic representation of the decision-making process.

2.5 Modelling Implications of the Choice Hierarchy

This hierarchical choice structure permits the development of two different models, one for mobility choices and one for travel choices. This set of models can be termed block-conditional (or in Ben-Akiva's original terminology block-recursive), where the blocks of mobility and travel choices as single units have a conditional (or a recursive) structure, while each block by itself has a joint structure.

This hierarchy, while theoretically pleasing, creates a very difficult modelling problem. The process determining the household's employment location is not well understood. Furthermore, for most workers,

*The utility functions can in general be parameterized in terms of the household's socioeconomic characteristics. Hence, the homogeneity of the utilities in a given market segment only implies that the same family of functions applies to each household.
employment location is relatively fixed. An employee must generally switch his employer in order to shift his job location. In order to avoid the complexities associated with modelling employment location decision, it was assumed that employment location choices are made on a much longer time frame than the remaining aspects of the mobility decision. This assumption, which is invoked in most empirical and theoretical studies of residential location, leads to a three stage hierarchy as depicted in Figure 2.2. The set of second stage choices including residential housing, location, auto ownership, and mode to work but not including employment location will be termed the mobility bundle.

This three stage hierarchy has important implications for developing a behavioral model of mobility choice. To consider these implications, it is useful to divide the entire set of factors which could possibly affect the choice of mobility bundle into four classes.

(1) Those entering into the decision process because they are the result of choices made "higher up" in the hierarchy (i.e. variables resulting from decisions about employment location)

(2) Those that directly affect the choice of the mobility bundle.

(3) Those which arise because of expectations from decisions "lower" in the hierarchy (i.e. variables describing decisions made conditional on the mobility choice).

(4) Those that directly affect the choice of mobility bundle as well as "higher" or "lower" level choices.
Figure 2.2

THREE STAGE HIERARCHY OF CHOICE
A brief example will make this distinction clearer. Consider the following four variables:

(1) whether a worker in the household is employed in the central business district (CBD);

(2) rent (or housing price);

(3) level of service by transit for shopping trips;

(4) level of service by transit to work.

The first of these variables, whether the workplace is in the CBD, is the result of the employment location in the hierarchy; hence it is predetermined when the household makes it mobility choice. It is an attribute of the household and does not vary among alternative mobility bundles. The second variable, rent, is an attribute of the alternative location and housing combinations; hence it has a different value for each alternative. The third variable, shopping level of service, depends on how the household will choose to travel for shopping, a decision which is made only conditional on the household's mobility choice. Thus, shopping level of service is said to be indeterminate in the mobility decision. This does not imply that it does not enter into the household's evaluation of mobility alternatives; rather, it means that the household does not know the specific shopping trips it will make. It does have some composite picture of the overall shopping level of service it will obtain for various mobility bundles.* For example,

---

*This concept of a composite measure for variables which are indeterminate has been utilized by CRA (1972) and Ben-Akiva (1973). The way in which variables representing these measures were formulated in this study is discussed at length in Chapter 7.
if the household were to choose to own two or more autos in its mobility bundle, it would generally have a high composite level of service for shopping associated with suburban locations because it would typically travel by car. Alternatively, if the household chose not to own an automobile, suburban locations would have poor composite level of service since it would have to shop exclusively by transit. The last type of variable, transit level of service for the work trip, is an attribute of an alternative mobility bundle as well as an attribute of alternative employment location choices.

For some households, the three stage choice hierarchy may not be representative of the mobility decision-making process. Households in which workers are chronically unemployed may change job locations very frequently. (Ingram, et. al. 1972). These households probably do not consider any single place as their employment location. The location choice process for these households is explored further in section 4.10. In the empirical study described in later chapters of this thesis, an attempt was made to eliminate these households in order to restrict the focus of the research to those households for whom the assumed choice hierarchy is relevant.

### 2.6 Dimensions of the Mobility Bundle

Given the structure of the choice hierarchy in Figure 2.2, the next question to consider is precisely what the mobility bundle consists of.
Location, housing, automobile ownership and mode of travel to work can be almost infinitely subdivided. Alternative locations can be taken to be cities, towns, census tracts, blocks zones, or any other geographical unit. Alternatively, location can be defined simply in terms of distance from the CBD or whether a location is in the central city, urban ring, suburbia, or rural fringe. Housing can be defined along a broad spectrum of dimensions, including age of structure, lot size, architectural style, number of rooms of various types, garage space, quality and condition of unit and type of tenure. Automobile ownership can consist of the number of autos as well as their make, age, gas mileage, horsepower, or operating cost. Finally, mode to work can be roughly classified as transit or car, or further described as bus, trolley, rail rapid transit, taxi, shared ride, paid car pool, or drive alone.

Clearly, at some level of detail the number of possible alternative mobility bundles one could create is enormous. Even if suitable data were available, a model developed with such detailed alternatives would be almost impossible to estimate and apply. Some level of abstraction in defining alternatives is clearly indicated.

In the case of the location dimension of the mobility choice, reducing the number of alternatives presents some basic methodological difficulties. Ultimately, each household selects a particular dwelling unit, of which there may be millions in any one metropolitan area. Any method of grouping alternatives is by its very nature somewhat arbitrary.
Census tracts, traffic zones, or other planning districts are truly arbitrary, since the decision-making unit most often doesn't even know where such geographical boundaries are. Towns or other jurisdictional boundaries are somewhat better, but frequently encompass many different neighborhoods with completely different characteristics.

This dilemma can fortunately be resolved. As will be shown in Chapter 5, it is possible to use certain choice theory approaches which rely on data about groups of dwelling units but provide reliable estimates of how households perceive the dwelling units themselves. This result allows the theoretical development to proceed with each separate dwelling unit as a distinct alternative without placing impossible computational and data requirements on the empirical study.

The problem of defining alternatives is much more difficult in the case of the housing choice. Housing is a ranked alternative, i.e. each different type of housing has a different set of identifiable characteristics which are not easily described by measurable attributes.* Location, on the other hand, is primarily an unranked alternative in that for the most part a location has very few intrinsic properties which make it unique. For this reason, it is quite difficult to restrict the dimensions of the housing decision.

---

*The definition of precisely what constitutes a ranked or unranked alternative is not absolute. Rather, it is a question of degree. For example modes such as car and bus are generally considered to be ranked alternatives. However, if one used a very detailed set of variables to describe each mode so that all the relevant attributes which distinguish car from bus were completely represented, it might be possible to view modes as unranked. See Ben-Akiva (1973) for a further discussion of this issue.
The dimensions of housing alternatives used in this study are structure type and tenure. This choice was determined primarily by the data available. For the purposes of transportation planning, where the primary focus is on the spatial aspect of mobility rather than housing, this choice set should prove adequate. However, the use of only structure type and tenure does restrict the applicability of this study to the analysis of housing policies, where issues of structure size and quality are very significant.

Automobile ownership is also a ranked set of alternatives. In addition, one aspect of the auto ownership choice, the number of autos a household selects, is logically ordered. (Ben-Akiva, 1973). The transportation planner is primarily interested in how people will alter their travel behavior in response to various policies. This response is primarily determined by the number of automobiles owned by the household, not the types of autos owned.* Thus, for the behavioral questions this study is intended to address and for the types of forecasting problems to which the models developed in this study will be applied, a consideration of the number of autos should be sufficient.

The final choice dimension, mode of travel, was restricted to two modes of vehicular travel, car and transit. These two modes together constitute 93% of all work trips in the study city, Washington, D.C. (This figure does not include some very short work trips). Thus, using

* An exception to this is when pollution control strategies are being evaluated. In this case the mix of cars as well as their number is critical.
only car and transit captures the choices of a vast majority of the population. In order to further limit the scope of the empirical study, only the auto driver mode was considered. All forms of ride sharing, some 18% of all car work trips in Washington, were eliminated.

Not all possible residential location/housing/auto ownership/mode to work combinations are necessarily feasible alternatives. For example, some locations are not served by transit, and therefore for households residing there using car to work is essential. Similarly, some locations have a limited range of housing options due to zoning ordinances. Furthermore, certain households may not have some mobility bundles open to them. Households without drivers obviously do not own automobiles, and low income households can only afford a limited subset of all feasible mobility bundles.

A detailed definition of the set of alternatives and the restrictions on which alternatives apply to each household are presented in Chapter 7.

2.7 Factors Affecting Mobility Decisions

Instead of reviewing the literature for the specific explanatory variables that have been used in the past for forecasting mobility choice, the focus in this section is on the actual underlying behavioral variables which should play a role in determining the mobility decision. This will hopefully yield insight into some of the inadequacies of existing methodologies (reviewed in Chapter 3) and provide a behavioral basis for selecting variables which should be considered in an empirical study.
The variables which affect the choice of mobility bundle can be divided into six general categories. These are as follows:

1. **transportation level of service to work** - travel time (in-vehicle and excess time) cost, comfort, convenience, etc., all for the work trip;

2. **automobile ownership attributes** - taxes, depreciation, registration costs, maintenance, title costs, etc.;

3. **locational attributes** - neighborhood quality, demographic composition, taxes, urban services, parking availability, local insurance rates, etc.

4. **housing attributes** - age of structure, quality, size of unit, garages, driveways, structure type, etc.;

5. **spatial opportunities** - measures of accessibility to shopping, social, school, personal business, and recreational destinations;

6. **socioeconomic characteristics** - income, race, household size, number of drivers, number of workers, education, marital status, age of household members, etc.;

The first type of variable, **transportation level of service to work**, can be measured for both the transit and car modes. Each potential residential location is characterized by a vector of level of service measures for travelling to the workplaces of household members. These values influence the choice of location and the choice of mode directly, but also influence the auto ownership and housing selection through the joint structure of the mobility decision. For example, if transit to work is quite favorable as compared with the car mode in a given location, the household choosing to live there might have a lower probability of owning many cars and living in a housing unit with a two car garage.
The next class of variables, auto ownership attributes, clearly impact directly on how households evaluate alternative levels of automobile ownership. Expenditures on automobiles, for example, may represent a substantial portion of a household's income, and are the primary reason why households frequently limit their auto ownership to something less than one per driver.

Locational attributes are variables which reflect the desirability of alternative residential sites. This category includes neighborhood characteristics such as crime rates, the physical appearance of the area, local taxes, and urban services such as police, garbage collection, and schools. The effect of a number of locational attributes on the mobility decision is explored in Chapter 4.

Housing attributes are closely tied to locational variables. While it is conceptually possible to separate housing from land costs (or rents), households make location decisions based on the combined cost. Housing can also influence auto ownership and mode to work. For example, it is a common hypothesis (which has yet to be convincingly verified empirically) that minority households, because they are denied access to higher quality housing by discrimination in the housing market, often substitute auto expenditures for housing expenditures.

The spatial opportunity variables reflect the transportation level of service for non-work travel. As discussed in section 2.5, these variables must be measured as some composite level of service because the
travel decisions are made conditional on the actual mobility bundle selected. Spatial opportunities, like other level of service measures, can be defined for both the car and transit mode.

Socioeconomic characteristics reflect the fact that even if the preferences of the members of a market segment are the same, there are certain characteristics of the group members which alter their opportunities and needs. For example, low income families may simply not be able to afford a single family dwelling and therefore do not perceive it as a relevant alternative. Also, the introduction of socioeconomic variables reflects the possibility that the members of a group may all have the same family of utility functions, where the actual function pertaining to any particular household depends on its specific socioeconomic characteristics.

2.8 Summary of Chapter 2

The general conceptual framework developed in this chapter relies on a choice theory approach to modelling urban mobility decisions. Households make their transportation-related decisions in a choice hierarchy in which the mobility choices of location, housing, automobile ownership and mode to work are made first, followed by the travel decisions of trip frequency, destination, mode, route and time of day for non-work trips, made conditional on the outcome of the mobility decision. This hierarchy was further simplified by separating the employment location
decision from the rest of the mobility choice. The remaining set of mobility alternatives was defined as the mobility bundle. It should be made clear that the choice hierarchy proposed in this chapter is a structural assumption which is not actually verified in the course of the study. Empirical studies can rarely demonstrate that a particular model form is the true one. Rather, the choice hierarchy is a working hypothesis which cannot be empirically rejected from the results of the models.

Six general classes of factors which potentially influence the household's choice of a mobility bundle were defined. These classes are transportation level of service to work, automobile ownership attributes, locational attributes, housing attributes, spatial opportunities, and the socioeconomic characteristics of the household. These different types of variables can influence the mobility decisions either directly or by affecting other decisions in the choice hierarchy.

The following chapter describes some of the previous research into mobility decisions. A broad spectrum of approaches, including sociological analyses, economic impact studies, and comprehensive urban land use models are considered. Each previous study is examined with particular emphasis on the behavioral choice structure which is either explicitly or implicitly used.
3.1 Structure of the Literature Review

The relationship between urban location patterns and public policy has been the subject of a large body of both theoretical and empirical investigation. Sociologists, psychologists, urban economists, urban planners and transportation systems analysts have all adopted different perspectives and methodologies in order to examine particular issues relating to various aspects of the urban mobility decision. This chapter reviews some of these past empirical and theoretical efforts. It is not intended to be an exhaustive review; instead, it attempts to examine a broad range of methodologies and perspectives to hopefully gain insights into how the individual household's mobility choice process should be represented in a behavioral model.

The first class of research efforts considered includes various theoretical models of the spatial structure of cities. This class includes simple descriptive models with limited behavioral content as well as more comprehensive theories of an urban system.

The following section reviews a number of sociological studies of urban location and mobility. These studies rely primarily on survey techniques which explore the preferences and attitudes of urban dwellers.

The third type of work examined includes empirical studies of urban phenomena which implicitly rely on reduced form models of a simultaneous supply/demand process to obtain information about the impact of certain exogenous variables on rents and prices. These efforts, termed here
impact oriented studies, include a large body of research, only a small fraction of which is considered.

The next type of research reviewed is the set of comprehensive urban land use models. Such models are typically extremely general in nature and are designed by economists, transportation analysts and urban planners to evaluate the impacts of alternative public policies on urban form. Because these models are frequently quite complex and have many submodels, each of which may be the result of a major empirical effort, particular emphasis is placed on the way in which the mobility choice is implicitly or explicitly treated. However, the general structure of each model is described in order to examine the way in which the submodels determining mobility decisions interact with the locational decisions being made by firms and other actors in the urban area.

The last type of research reviewed includes efforts to estimate behavioral models of various aspects of the mobility bundle. Particular emphasis is placed on models which are developed from a household level theory of mobility choice.

The descriptions of these prior studies are followed by a discussion of their major implications for modelling the mobility choice process as outlined in Chapter 2.

3.2 Theoretical Studies of Urban Mobility

Some of the earliest theoretical analyses of urban location focused on describing the spatial pattern of urban growth. The models which resulted from this work are almost entirely descriptive in nature and
rely primarily on qualitative descriptions of the location process rather than on specific analytic models.

Two general theories have become established in this field. The first, which might be termed the "concentric ring" hypothesis, describes urban growth in terms of the radial movement of activities from the urban center. This theory was first developed by Hurd (1903) based on observations of the growth of a number of major U.S. cities. An alternative approach, termed the "sector model," was developed by Hoyt (1939). In this model, as cities grow activities move outward in particular directions rather than uniformly in all directions. The result of Hoyt's model is a city in which activities are distributed along corridors, or sectors, rather than in concentric rings.

Both Hurd and Hoyt recognize the basic importance of transportation facilities in determining how cities grow. In both models, residential growth occurs at the city periphery. Hoyt points out that the best residential areas tend to develop along the lines of best access in the city.

A model which is a fusion of the sector and ring theories was developed by Hawkes (1967). In Hawkes' model, the city grows both axially and radially along major transportation lines. Using census tract data for a number of cities, Hawkes calibrates polar coordinate models for the distribution of various activities. These models have as independent variables only the distance from the CBD and angular orientation of the tract with respect to an arbitrary base line, and therefore offer virtually no insight into the causal mechanisms underlying city formation.
An alternative approach to how cities grow is postulated by Verona (1973). He hypothesizes that the way cities are formed is determined by an upper bound on the travel time to the city center. This limit, which he places at 45 minutes, represents the greatest amount of time a person will typically devote to travelling to work. As transportation technology improves, the area which can be spanned in 45 minutes grows; hence, outward development occurs. Travel corridors along major arterials expand the city in their direction. Verona considers the earliest cities from this perspective, and shows that as travel speeds increased the maximum travel times remained relatively constant.

While greatly oversimplified, Verona's hypothesis may have some empirical foundation. Analysis of travel time budgets by Zahavi (1974) confirms that households devote a fairly constant amount of time to travel activities. However, Verona's theory fails to explain why some cities grow to be larger than others even though they have the same available transportation technology. More significantly, Verona's approach provides no insight into the way in which activities will be distributed over the city.

The above models are all extremely naive representations of the decision process which give rise to urban location patterns. Recent efforts have attempted to build more detailed behavioral theories of spatial structure from economic theory. These theories can be characterized as either aggregate or disaggregate, depending on whether the underlying behavioral processes are described as pertaining to a group of decision-making units or each decision-maker separately.
The classic example of an aggregate theory is a model developed by Mills (1967).* In Mills' model, the city is viewed as a center of production which has inputs of land, labor, and capital. The city has as its sole outputs transportation, housing, and a composite good, which represents all other production. Mills' city is monocentric, and resources are used in the production process as labor is moved from residence areas to the central workplace.

The central business district in the model is the locus of all composite good production. The size of the area it occupies is determined by the area over which the price industry is willing to bid for land exceeds the residential rents. Some of the important conditions which must be maintained for a city to be at equilibrium are as follows:

1. Production of composite good = \( f_1 \) (land, labor, capital).
2. Land in transportation = \( f_2 \) (transportation service provided).
3. Housing = \( f_3 \) (land, labor, capital).
4. Wage and interest fixed and exogenously determined.
5. Demand for composite good = \( f_4 \) (prices of composite good).
6. Price of land at fringe equals prevailing agricultural prices.
7. Land used for production of composite good + land used for transportation exhausts CBD land.
8. Land used for housing + land used for transportation exhausts non-CBD urban land.
9. Production of composite good and housing is performed with optimal resource allocation at all locations in city.

Conditions (1) through (3) and (9) are all defined so as to pertain

---

*Other such models include those of Solow and Vickrey (1971), Rubinfeld (1972), and Yellin (1974).
at all points in the city. Condition 9 insures that the mix of resources
used in production varies over the city, since the cost of land varies
over the city. With appropriate assumptions about the form of the pro-
duction functions $f_1(\cdot), f_2(\cdot), \text{ and } f_3(\cdot)$ and the demand function
$f_4(\cdot)$, Mills solves for the rent surface, the boundary of the CBD, the
size of city, and the resources used in the city.

For the purposes of this study, a more useful approach is charac-
terized by the models of Wingo (1961), Alonso (1964), and Muth (1969)*.
While these models are all somewhat different, each begins with a simpli-
fied household level utility function and budget constraint which express
the household's preferences for aspects of its location and other
consumer goods. As in classical consumer theory, every household seeks
to maximize its utility subject to a constraint imposed by its disposable
income. Firms in the urban area attempt to maximize their profits and
compete for space with households and each other in an open land market.
The pattern of location and land prices in the city is the outcome of a
market equilibration process.

The set of conditions which must be satisfied for a city to be in
equilibrium can be described as follows:

(1) Each household maximizes its utility subject to a budget
constraint.

---

*One portion of Alonso's model forms the basis for a behavioral analysis
of household location in the next chapter. For this reason, a discussion
of the details of these models is omitted here.
(2) Each firm maximizes its profits, which must be non-negative.*

(3) The land price at the edge of the city equals the prevailing agricultural price.

(4) All urban land is exhausted.

(5) All markets behave competitively with perfect information available to all firms and households.

Both the aggregate and disaggregate models discussed above and later extensions to them have a number of similar characteristics. First, they are monocentric; all significant trip-making is directed towards a central point. As will be considered in more detail in Chapter 4, this perspective seems somewhat unreasonable in light of the post-World War II trend towards the decentralization of employment in U.S. cities. A second characteristic of these models are that they are static in nature. The city is assumed to be formed in one instantaneous step, rather than incrementally over time. Finally, the models assume that the city is in equilibrium. This requires that the market resolution process is completely frictionless and that all actors in the process have perfect information about the land market.

These limitations of theoretical models of urban form, as well as some others described in the next chapter, have greatly limited their usefulness as practical forecasting tools. Even the strongest advocates of theoretical models would probably agree that their most important

*Since all firms are manufacturing a single composite good with the same production function, the profit all firms make in the urban area are equal. In the long run, these profits should be zero; however, this is not essential to the theory.
contribution has been in their formalization of basic concepts about the nature of the spatial distribution of activities in a city rather than their direct applicability to problems of policy analysis.

3.3 Sociological Studies of Urban Location

To the social scientist, the concentration of activities which exists in urban areas and the very high level of interpersonal interaction which results provide the perfect framework for the study of social systems. It is for this reason that the city has always been the focus of a great deal of sociological investigation. Any attempt to exhaustively consider this large body of research within the framework of this thesis would be doomed to failure. However, a few studies have been oriented towards considering how households make their mobility decisions and how they perceive various aspects of the mobility bundle. It is these studies which merit careful examination here.

Such studies have typically relied on surveys of households to draw inferences about mobility decisions. These surveys are generally used in two ways. First, responses are tabulated to draw inferences about the distribution of reported perceptions and attitudes. Second, non-quantitative responses to questions are frequently recorded and used in an anecdotal fashion to illustrate or confirm hypotheses. Both approaches, if carefully used, can be effective analysis tools. In general, survey data is rarely used in these studies in an explicit statistical modelling procedure.

In a study relying on the tabulation of data by Lansing and Mueller (1966), a national sample of urban households was questioned about their
locational and transportation preferences.* This data set was collected as part of an ongoing program of the Survey Research Center at the University of Michigan. The results of the study indicated the following:

(1) Households are highly mobile. Over one-half (54%) of the respondents indicated that they would move in the next five years.

(2) Households consider a broad range of factors in making their mobility decisions. Those who moved in the last five years were asked why they considered their location choice a good or bad idea. They listed the following factors:

<table>
<thead>
<tr>
<th>Factor</th>
<th>% Responded**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing unit itself</td>
<td>32%</td>
</tr>
<tr>
<td>Neighborhood considerations</td>
<td>22%</td>
</tr>
<tr>
<td>Non-work locational factors</td>
<td>18%</td>
</tr>
<tr>
<td>Cost considerations</td>
<td>14%</td>
</tr>
<tr>
<td>Other</td>
<td>14%</td>
</tr>
<tr>
<td>Proximity of Workplace</td>
<td>13%</td>
</tr>
</tbody>
</table>

Note that proximity to workplace is the last of the factors listed. This has significant implications for both theoretical and comprehensive models which attempt to explain location only in terms of work trip level of service.

(3) In general, households have a preference for the life style associated with single family homes further from the CBD. An exception to this is families with very low incomes, who indicate a preference for living near the city center.

---

*New York City residents were not included in this sample.

**Respondents were allowed to choose two factors. Hence, percentages do not sum to 100%.
(4) Neighborhoods are perceived in terms of a number of attributes, the most important of which are the types of people living there, the area's physical characteristics, and the convenience of the location for reaching non-work destinations.

Lansing and Hendricks (1967) use the same data source to explore how two household mobility choices, residential density and automobile ownership, are determined. As in the previously cited study, respondents were asked about the decisions they made, the reasons they made them, and what alternatives they would prefer. As before, households in general indicated strong preferences for suburban style living, as represented by the single family home. The percentage expressing this preference varied by life cycle from 96% for married couples with children to 54% for young single persons.

A number of studies which rely more heavily on the anecdotal aspects of survey responses are also of relevance. Coleman (1973) used surveys from Boston to investigate how housing alternatives are perceived. He argues that housing is viewed by people as existing in seven distinct levels as follows:

(1) Prestige Class – estates, mansions and other luxury dwellings (estimated 1.6% of Boston stock);

(2) Very Good – large colonial or custom-built contemporary homes (estimated 4.9% of Boston stock);

(3) Pleasantly Good – Capes, split-levels or ranches of moderate size in good condition (estimated 15.1% of Boston stock);

(4) Standard/comfortable – post-war tract houses or pre-war Capes (estimated 25.4% of Boston stock);

(5) Standard/marginal – smaller units in multi-family structures, somewhat less desirable than type 4 (estimated 26.9% of Boston stock).
(6) Substandard housing in projects or older structures below inspection standards but rehabilitable (estimated 21.5% of Boston stock);

(7) Slum units already abandoned or should be (estimated 4.6% of Boston stock).

Coleman defines each level as having associated with it a neighborhood quality and a type of residential population. For example, he defines "very good" housing as associated with professional and managerial families with annual incomes in the range from $22,400 to $34,900.

This type of analysis has also been applied by Anderson-Khleif and Coleman (1974) to explore public attitudes towards new housing forms. Included in this study of 300 Kansas City residents was an examination of the acceptability of condominiums, pre-fabricated housing, and mobile homes. Attitudinal responses were distinguished by the social status of the respondents (upper-status, middle class, and working class). The results indicate a substantial amount of variation in the preferences across these groups. As one might expect, lower status households tended to view such innovations much more favorably than did upper status respondents.

The major limitation of this entire area of research is that it is often impossible to separate people's responses from their economic condition and perceptions. Asking people what their preferences are over one attribute of housing without precisely defining the other attributes results in respondents applying their own perceptions of what the other attributes might be. For example, in the Survey Research Center survey, respondents were asked the following question:
"If you could do as you please, would you like to live closer to the center of (....METRO AREA....) or farther from the center of (....METRO AREA....) or just where you are?" (Lansing and Hendricks, 1967).

In answering this type of question, the respondent supplies his own image of the type of community each location might have, the type of housing he would live in, and how much it would cost him. The response perhaps provides insight into questions which are of interest to the sociologist such as where he would live given his current images of the alternatives presented to him. However, it doesn't answer the following question:

"If given a choice between two identical units (i.e. identical structures, prices, and neighborhoods), one closer to the city, and one further away, which would you choose?"

Survey techniques which assess preferences using the latter type of question are currently the object of fairly intensive research (Hauser, 1974). Such techniques, termed direct utility assessment, may eventually provide information which is of great use to the transportation planner by allowing an in-depth assessment of the tradeoffs people make in determining their mobility choices.

3.4 Impact Oriented Studies

Urban economists have frequently postulated that a variety of specific causal factors enter into a household's perception of location and housing alternatives. An entire class of models, termed here for convenience impact oriented models, have been estimated to test these
hypotheses.

These studies all tend to have a common theoretical basis.* Households are assumed to demand location/housing packages in certain quantities according to a demand function as follows:

\[ Q = D(P, X_1, X_2) \]

where:

- \( Q \) is the quantity demanded;
- \( P \) is some price or value variable, and,
- \( X_1, X_2 \) are vectors of exogenous variables.

Similarly, as in classical economic theory, entrepreneurs determine the prices for housing in various locations according to some supply function as follows:

\[ P = S(Q, X_1, X_3) \]

where:

- \( P, Q, X_1 \) are defined above, and \( X_3 \) is another vector of exogenous variables.

Using the two equations, it is possible to derive a reduced form equation, denoted as follows:

\[ P = g(X_1, X_2, X_3) \]

in which prices are determined entirely as a function of exogenous variables. It is this equation that most impact oriented studies have typically attempted to estimate. Generally, the function \( g(\ ) \) is defined as linear in its parameters and ordinary least squares methods are used. One output of these models is a set of estimates of the

*In many cases, this basis is implicit rather than explicit.
implied market value of various housing attributes such as an additional room, a unit of lot size, or various public services. These implicit market evaluations are generally termed "hedonic price indices," and can potentially be used for tax valuation purposes or to make market value forecasts.

Another smaller group of studies has tried to isolate the effect of one exogenous variable yet avoid the need to actually estimate the reduced form. These studies use either "before-after" data or cross-sectional comparisons of nearly identical units. This group of studies relies on the assumption that the only factor which differs among comparable mobility bundles is the variable of interest, and its impact can therefore be isolated without correcting for the effect of changes in other variables.

Each study is generally an attempt to demonstrate the importance of a specific exogenous variable or group of variables. However, mis-specification of the reduced form due to the omission of significant variables can result in misleading conclusions about the effect of the included variables. For this reason, this class of studies is generally characterized by very large numbers of independent variables.

Table 3.1 presents a summary of a number of impact oriented modelling studies. In each case, the researcher, the year of the study, the primary data sources, the type of model form, the dependent and the exogenous variables used, and the variables which were highlighted as important in the study are listed. Some of these studies are considered further in the next chapter as a basis for introducing certain factors
<table>
<thead>
<tr>
<th>Researcher</th>
<th>Year</th>
<th>Data Sources</th>
<th>Type of Model</th>
<th>Dependent Variables</th>
<th>Exogenous Variables</th>
<th>Highlighted Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brigham</td>
<td>1965</td>
<td>Los Angeles tax appraisals (cross-sectional)</td>
<td>linear regression</td>
<td>site value for residential, commercial and industrial land</td>
<td>accessibility; neighborhood quality (amenities); topography</td>
<td>location of amenities</td>
</tr>
<tr>
<td>Ridker &amp; Henning</td>
<td>1967</td>
<td>St. Louis (cross-sectional)</td>
<td>linear regression</td>
<td>value of single family housing</td>
<td>housing unit characteristics; accessibility; neighborhood quality; taxes and public services; racial composition; pollution (sulfation levels)</td>
<td>pollution</td>
</tr>
<tr>
<td>Oates</td>
<td>1969</td>
<td>northeastern New Jersey (cross-sectional)</td>
<td>linear regression &amp; two stage least squares</td>
<td>value of single family housing</td>
<td>housing unit characteristics; accessibility; taxes; school expenditures</td>
<td>school expenditures</td>
</tr>
<tr>
<td>Massel &amp; Stewart</td>
<td>1971</td>
<td>Palo Alto, California (cross-sectional)</td>
<td>multiplicative form</td>
<td>value of single family housing</td>
<td>housing unit characteristics; neighborhood quality; lot size; zoning</td>
<td>zoning and neighborhood quality</td>
</tr>
<tr>
<td>Mudge</td>
<td>1973</td>
<td>Lindenwold High Speed Line Corridor tax data &amp; census</td>
<td>linear regression</td>
<td>sales prices</td>
<td>housing unit charac.; neighborhood quality; lot size; accessibility; time savings due to transit implementation</td>
<td>accessibility (by transit)</td>
</tr>
</tbody>
</table>

**TABLE 3.1 - IMPACT ORIENTED ECONOMETRIC MODELS**
into the household's evaluation of alternative mobility bundles.

The measures used for the exogenous variables vary widely from study to study. Oates, for example, describes accessibility only in terms of the proximity of a community to the central business district (in his study, Manhattan). Alternatively, Brigham uses a weighted function of employment defined as follows:

$$\text{Accessibility of zone } i = \sum_{\text{all workplaces } j} \left( \frac{\text{employment}_j^c}{a + b(\text{distance})_{ij}} \right)$$

where $a$, $b$, and $c$ are constants. Similarly, the level of aggregation used in the data also varies. Oates used community level census data, while Mudge used the actual sales prices of individual units. Both Massel and Stewart and Mudge applied factor analytic techniques to reduce the number of variables which describe structure types and neighborhoods.

The only study which attempts some consideration of the problems of simultaneity between some of the exogenous variables and prices is Oates'. He explicitly recognizes that tax rates are probably a function of home values, since communities make their taxing decisions based partly on the value of taxable property. He therefore uses two stage least squares, a simultaneous equation estimation procedure, to eliminate a potential source of inconsistency in the parameter estimates.

The other group of impact oriented studies, those relying on a comparison between nearly identical units, has often focused on the effect of racial integration on property values. Laurenti (1972) used cross-sectional data from recently sold houses in Philadelphia, San
Francisco and Oakland. He attempted to isolate particular units with identical characteristics but located in areas with different racial compositions. After adjusting prices for differences in the way in which the units were financed, he concluded that non-whites pay more for equivalent housing than did whites, and that prices paid for housing in areas undergoing racial change were if anything slightly enhanced, presumably by the increased demand of non-whites who previously did not view the neighborhood as a feasible residential location. Ladd (1962) addresses the same question using time series data on assessed values from 1950 to 1960 in Ann Arbor, Michigan. He constructed three groups of houses, all of which were sold twice in the period of interest:

1. those sold both before and after racial integration;
2. those sold twice in an all white area;
3. those sold twice in a racially integrated area.

Unlike Laurenti, Ladd concluded that there are no significant inter-group differences in prices based on the neighborhood's racial composition.

Both the econometric studies and the comparative analyses discussed above suffer from a common defect. By considering only the reduced form rather than the structural equations, it is impossible to unravel the demand and supply effects of different variables. A small or statistically insignificant coefficient in the reduced form does not necessarily mean that the corresponding variable is not very important in determining supply, demand, or both. In short, the studies can only provide insight into the total net market effect; they provide no
specific information about the nature of the demand function. In fact, a very similar approach has been taken by other researchers such as Rabe and Hudson (1974), in which the effect of specific variables has been related to the supply of housing.*

This difficulty is compounded by the previously mentioned problem of potential mis-specification. There always remains the possibility that the variable of interest is in some way correlated with an omitted variable, and hence the estimated coefficient is measuring more than one impact. For this reason, these studies are at best only indicators of factors which might appear in models of structural, or behavioral, equations. In the final analysis, some form of the structural demand equation must be estimated before conclusions about what factors affect demand can be reached.

3.5 Comprehensive Urban Land Use Models

Unlike urban economists or sociologists who are often concerned solely with understanding urban phenomena, transportation planners are most frequently concerned with forecasting future patterns of land use. This orientation has lead to the formulation of a number of comprehensive urban land use models.

Each of these models usually consists of a number of submodels to forecast the location of various activities. In this review, only the

---

*Rabe and Hudson focus on the quantity reduced form equation rather than the price equation. However, their results still do not produce estimates of the actual structural coefficients. Unlike many of the above studies, they explicitly recognize this shortcoming.
residential location and housing components of these models will be considered, since these submodels have embedded within them assumptions about the nature of the household's mobility decision. Furthermore, models which were proposed but never actually implemented, such as the Penn-Jersey model based on research into optimal location patterns by Herbert and Stevens (1970), are not considered.

For the purposes of this review, urban land use models are divided into three categories:

1. Econometric models which use explicit functional representations of urban behavior and econometric methods for parameter estimation;
2. Lowry-based models which are derived from the original work of Ira S. Lowry and therefore share a common set of assumptions;
3. Urban Simulation models, a separate category for the NBER-Urban Simulation model.

Each of these model types is considered in a subsection below. Also, a review of perhaps the most cogent critique of comprehensive land use models by Douglas Lee is discussed.

- Econometric models

The earliest member of the first group of models is EMPIRIC, developed in by Traffic Research Corporation (1966) for the Boston metropolitan area.* In this model, the total population in each of a number of socioeconomic groups is assumed to be exogenously forecast. This population (as well as a number of employment activities) is allocated

*Another model developed in this study, POLIMETRIC, is very similar in structure to EMPIRIC and is therefore not examined in this review.
to zones by determining the change in the share of the control total which will occur in every zone over each time period. Exogenous variables include zone accessibility, water service and sewage service. In EMPIRIC, each population and employment activity is allocated by a single equation which is estimated and applied simultaneously with the other activity location equations. The structure of the population allocation equations implies the following assumptions.

(1) Households make location decisions with reference only to a general accessibility rather than a known employment location as assumed in the choice hierarchy in Figure 2.2.

(2) The housing stock is sufficiently uniform over the city so that it can meaningfully be ignored in the location decision.

(3) Prices for locations do not significantly determine spatial preferences.

These assumptions have given rise to serious doubts as to the behavioral content of the EMPIRIC model. Factors such as neighborhood quality, schools, taxes, and race which previous studies have indicated as being significant factors in location decisions are entirely ignored. Furthermore, supply and demand effects seem to be confused in the structure of the equations. For example, the installation of sewerage facilities might more reasonably be treated in an equation describing the supply of housing provided by developers rather than as part of the household's location decision.

EMPIRIC does explicitly treat the effect of the demographic composition of a zone on its desirability as a location alternative. The variables representing the number or change in the members of one socio-economic group often enter into the equations describing the change in
share of another group. For example, the change in the share of high income residents in a zone might be decreased by the number of low income households present. Furthermore, EMPIRIC is a dynamic model; forecasts in each period are affected by prior period forecasts. This is done by introducing time-lagged dependent variables into the structural equations.

Many of the difficulties in EMPIRIC have been corrected in an econometric model still being developed by Bradbury et al. (1974). In their demand equation, the demand for housing is modelled as a function of prices, housing type, the age of housing stock, residential density, neighborhood quality and racial composition, taxes, the amount of available recreational space, school quality, and accessibility by both highway and transit to employment. This model has been estimated using data at the town and neighborhood level in Boston. As in EMPIRIC, aggregate data is used and workplaces are treated as indeterminate. Furthermore, many of the coefficients estimated have very high standard errors, and are therefore not significantly different from zero at reasonable levels of confidence. This probably results from the high level of spatial aggregation used in the analysis.

b- Lowry-based models

The second class of models, those based on the original Lowry (1966) research, view the city from a very different perspective. All employment is divided into two major categories, basic and population-serving. Basic employment is defined as all employment activity the location of which is not affected by the proximity of residential population; this
is located exogenously. Population-serving employment is all remaining employment, and is divided into a number of classes depending on the variant of the model used.

The model iteratively seeks a pattern of location which satisfies the following conditions

(1) Total population-serving employment is a function of total basic employment.

(2) Total population is a function of total employment.

(3) The work trip length distribution is a function of distance or time from the workplace.

(4) The non-work trip (i.e., travel to population-serving employment) is a function of distance from the employment.

(5) Employees use services in their work zone.

(6) Zone employment must exceed a threshold for each employment type.

(7) There is a maximum residential density.

This general approach has evolved into a great number of land use models. (Goldner, 1971) This group of models, or "family tree" of the Lowry model, is illustrated in Figure 3.1.

The key assumptions about mobility decisions in the Lowry type models are embodied in the formulation of equation 3 above. The distribution of employees working in any single employment location is determined solely by some measure of the separation of that employment center from various residential locations. The functional form used to represent this relationship has typically been the gravity or intervening opportunity model.

As in EMPIRIC, prices and housing type are not incorporated into the
Figure 3.1
THE LOWRY MODEL FAMILY TREE
location model. Furthermore, unlike EMPIRIC, the demographic composition of a zone has no impact on its desirability as a residential location. The Lowry model also ignores the important array of other causal locational factors such as schools, taxes, public services, and neighborhood effects.

The Lowry model does, however, deal explicitly with a known workplace in the residential location submodel. Rather than using a weighted accessibility to employment, the actual work trip characteristics are used to determine the relative desirability of alternative locations.* Thus, unlike the EMPIRIC model, Lowry-based models are capable of producing a complete work trip origin-destination matrix which can be used directly in a modal split model. Putnam (1973) has done this with a Lowry type model calibrated for San Francisco with relatively encouraging results.

c. the NBER Urban Simulation Model

The last model to be considered is the National Bureau of Economic Research (NBER) Urban Simulation model (Ingram, et al., 1972). This model has a strong urban economic orientation and relies on an explicit land and housing market model of the city. Only one half of the entire model system, the demand side, will be considered here.

The demand for housing and location in the NBER model is represented by a series of submodels organized in a way which is analagous to the hierarchy of choice proposed in Chapter 2. In this sequence depicted in Figure 3.2, employment is first located by employment type in each

*In addition, the location of population-serving employment is determined by the location of the residential population.
Figure 3.2

THE MOBILITY SEQUENCE IN THE NBER URBAN SIMULATION MODEL
workplace. In the earliest versions of the NBER model, this function is performed as entirely exogenous. Work is now underway to introduce a population-serving employment location submodel into the overall model structure.

The next aspect of demand determined in the model is the household's decision to move. In each time period (usually a year) only a small fraction of the total population actually relocates. At present, the "movers" submodel simply applies socioeconomic group-specific mobility rates to determine the probability of any particular household relocating.

The third mobility decision in the sequence is the housing choice. Using housing prices from previous time periods aggregated across alternative locations, the members of each socioeconomic group are allocated to various housing types.* The earliest prototype model used linear regression equations to perform this allocation. The newer version of the model now uses the multinomial logit model. This submodel, developed by Quigley (1973), is discussed in greater detail in the following section.

The final step in the sequence is the resolution of the location market for each housing type. This allocation is performed by a linear programming model for each housing type in which the total system-wide travel costs are minimized. This approach is a variant of another linear

---

*The method by which the composite price of housing is determined includes travel costs. The composition rule has altered in the course of the model's development from a weighted averaging approach to the selection of the average price of the fifth percentile.
programming allocation model developed but never fully implemented by Herbert and Stevens (1970). It is based on a classical economic theory competitive market in which optimum resource allocation is achieved when each household maximizes its own benefit.

The results of the market clearing process produce "shadow prices" for each housing type in every zone, which are used in the next forecasting period in the housing demand and supply submodels.

\textbf{d - the Lee critique}

In a somewhat whimsical but important critique, Lee (1973) accuses comprehensive land use models of being guilty of "seven deadly sins," all of which collectively point to the need for a different emphasis in empirical and theoretical research. These "sins" are as follows:

(1) hypercomprehensiveness - the attempt to model all impacts and effects in a single model;

(2) grossness - the high level of spatial and demographic aggregation;

(3) hungriness - the need for enormous amounts of data;

(4) wrongheadedness - the difference between assumed causality and the actual structure of the model;

(5) complicatedness - the large number of variables and their possible interactions;

(6) mechanicalness - the use of approaches which computers are best suited to solve rather than those which are most causal;

(7) expensiveness - the enormous computational and manpower requirements of existing models.

Each of these criticisms is to some extent true for every model; however, each is also to some extent an overstatement. For example, it is indeed the case that land use models are very complicated.
Unfortunately, the system being modelled is also complex. Similarly, the data hungeriness of comprehensive models results more from the size and complexity of the system, as well as the need to avoid another deadly "sin" of grossness, than from any basic flaw in the structure of the models.

The more significant aspects of the Lee critique relate to the wrongheadedness and mechanicalness of existing models. If comprehensive land use models are to provide useful answers to policy questions, both these flaws must ultimately be eliminated. The only feasible way to do this is to draw on credible theories of household mobility behavior rather than exploring simple correlations among variables.

3.6 Behavioral Models of Mobility Decisions

There have been numerous efforts to directly model various aspects of the mobility decision. In order to focus on prior research which is of greatest relevance to this study, only models which rely on an explicit disaggregate, behavioral perspective will be considered. Most of these studies actually use disaggregate, i.e. household level, data; some resort to aggregate data but rely on a disaggregate theoretical development.

In order to further restrict the scope of this review, models which deal only with the choice of mode to work will not be considered. Such models are reviewed extensively in Ben-Akiva (1973) and Charles River Associates (1972).

Existing mobility choice models can be classified in a great number
of ways. The typology adopted here focuses on the form of the model structure used. Three basic structures can be defined:

(1) unidimensional – models which explicitly deal with only one aspect of the mobility bundle such as automobile ownership or housing;

(2) sequential – models which treat the aspects of the mobility decision in a sequence of choices, e.g. location, then housing, then auto ownership;

(3) joint – models which treat at least two aspects of the mobility bundle as determined by a joint, or simultaneous, decision process.

... unidimensional choice models

Theories of urban location such as those of Alonso, Muth, and Wingo predict that within a single socioeconomic group all households will locate within a ring at the same distance from a central workplace. In real cities, however, even if employees working outside the center city are ignored, socioeconomic groups tend to locate in overlapping rings. In order to deal with this, Mayo (1971) formulated a unidimensional probabilistic location model. This model is quite similar in its development to the disaggregate choice models to be discussed in Chapter 5. However, rather than focusing on the choice of location as a dependent variable, Mayo estimates the bid rent curves, i.e. functions representing the greatest amount households would be willing to pay for any location if their level of utility were held constant.* In Mayo's formulation, these bid rent functions are treated as being subject to random variation.

Coefficient estimates for his model were estimated using aggregate

---

*The concept of bid rents is treated more formally in Chapter 4.
data from Milwaukee. Locations were described by variables representing the demographic characteristics of the residents, the housing characteristics of the locations, the quality of public services, education, taxes, land use characteristics, shopping accessibility, and location rents. The population was segmented by income, race, the sex of the head of the household, household size, and the number of workers. Factor analysis was used to combine 35 possible variables into their five principal components in order to reduce the number of demographic and housing variables. Using the proportion of households in any socioeconomic group in each location as a measure of the probability of a member of that group selecting each location, Mayo estimated as series of linear probability models which reflect the probability a given location will have a higher bid rent than all others.

In a study with similar objectives but a somewhat different methodology, Wheaton (1974) estimates a series of log-linear and non-linear models of bid rents using aggregate data from San Francisco. Only workers in the central business district are considered. Rather than using Mayo's probabilistic model of bid rents, Wheaton relies on the theory that at equilibrium, all households in the same socioeconomic group will be on the same bid rent curve, or, more simply, households with identical bid rent curves will be equally well off. This assumption, along with some additional hypotheses about the additive separability of the utility function, permits the estimation of the relative level of utility for each socioeconomic group as the constant term in the bid rent curve. This is achieved by a theoretical derivation which briefly
summarized proceeds as follows:

The household's level of utility (after some derivation) can be written as

$$\text{Utility} = f_1(\text{location attributes}) + f_2(\text{income remaining after location expenses}) + \epsilon$$

where $\epsilon$ is a random element attributable to imperfect information and frictional factors in the location market. With appropriate assumptions about $f_1(\ )$ and $f_2(\ )$, one can estimate the function

$$f_2(\text{income remaining after location expenses}) = \text{Utility} - f_1(\text{location attributes}) - \epsilon.$$

Wheaton estimates this model for seven household strata using travel time to work, lot size, housing unit size, age of structures, average community income, and a measure of accessibility to shopping and other service facilities as independent variables. The last variable consistently had an unexpected negative coefficient, which Wheaton attributes to the fact that this variable may measure possible externalities associated with high accessibility such as noise, traffic and pollution.

In a model of rental housing choice, Quigley (1973) uses the multinomial logit model, which is described in detail in Chapter 5. Using household level data from the 1967 Pittsburgh Metropolitan Home Interview Survey, he considered the household's choice among three structure types: single family detached units, common wall units, and multi-family units; two age categories: pre- and post-1930; and three size categories: less than two bedrooms, two bedrooms, and more than two bedrooms. For each household, the total price of housing at every
location is assumed to be the sum of the contract rent, the value of the travel time needed to commute to work and the commuting cost. It is assumed that the minimum cost mode (where time costs are included) is always used and that the value of time is equal to the marginal wage rate. Using the surface of the total prices for each household, the actual cost of each housing type is assumed to be the average price of the least expensive five per cent of the existing stock of that housing type.

As independent variables, Quigley uses constant terms for common wall and apartment units (single family units are used as a base against which the relative desirability of the other two categories is measured), the number of bedrooms, an age of unit dummy variable, prices and the number of units of the housing type which exist in the metropolitan area. Thirty models for various socioeconomic groups defined by combinations of income and family size were estimated using maximum likelihood methods.

Ellis (1966) uses multi-class discriminant analysis to model housing choice for households with CBD workers.* Using census and home interview survey data from Tucson, Arizona, he defines groups of "environments," or housing bundles with attributes which seem logically related in the view of household's choosing housing. These environments are used as categories of housing in the model, and are defined along the following dimensions:

1. socioeconomic status factor - housing and neighborhood quality;
2. proportion of single family dwelling units;

*Actually, Ellis also develops a location choice model. However, this model is a simplistic intervening opportunities form which is neither disaggregate nor behavioral.
(3) recreational facilities
(4) racial composition;
(5) population density;
(6) age of housing;
(7) proportion of open space.

These dimensions result from an interpretation of the seven principal factors of eighteen distinct tract level variables. Each is then categorized into a number of levels such as high, middle and low social status or single vs. multi-family. These levels form the groups used in the discriminant analysis. In Ellis' model each dimension, or attribute, is treated independently rather than jointly. This greatly limits the credibility of his analysis since, for example, the event that a given household is classified by the model into the low socioeconomic status housing is certainly not independent of it also residing in a multi-family unit.

The variables used in the discriminant function are all descriptors of the household. Included are auto ownership (as a proxy for income), household size variables (number of members, number employed, percentage employed), trip-making variables (total auto trips, total trips per person, trips per person five and older), race, and the occupation of the head (blue collar or white collar). This extensive list of variables, which are at best instruments for true causal variables, makes the interpretation of the behavioral implications of Ellis' results somewhat difficult.

There have also been a number of unidimensional models of automobile
ownership developed. Burns, Golob, and Nicolaidis (1975) derive a logit model of auto ownership from a set of somewhat over-simplified hypotheses about how members of a household use automobiles. This theory requires that each auto a household owns is allocated solely to one member of the household, thereby making the complex interactions among household members analytically tractable.

The model, which is ultimately estimated using survey data from Detroit, includes variables representing the income remaining to the household after auto expenditures and the level of accessibility obtained by the household members, as well as constant terms for two of the three auto ownership levels.

Hoxie (1970) used discriminant analysis to estimate an auto ownership model with data from the Boston 1963 Home Interview Survey. He used a dummy variable for the household's life cycle and included insurance costs, which exhibit substantial variations across towns in Boston, as an independent variable. However, due to the unavailability of good level of service data, Hoxie was forced to measure transportation level of service by the average walk time to the transit system from the household's zone of residence.

**b - sequential models**

Kain (1964) estimated a series of aggregate models which assume a sequence of household decisions as follows: housing type, car ownership, mode to work, and finally distance of residence from work, which is a proxy for residential location. While this model is aggregate in orientation, it is discussed here because it is the first model found
which treats all the aspects of the mobility bundle in a single model structure.

Kain only considered the decisions of white households with one worker. Using data from a 1953 survey in Detroit, four aggregate linear equations were estimated. Independent variables include demographic descriptors of locations (mean family income, fraction of white male workers, labor force participation, and family size) and the level of service variables (distance from the CBD and a crude measure of transit service).

Using discriminant analysis, Aldana (1971) estimated a series of disaggregate models which are based on the assumption that households first select their location, then their auto ownership, and finally their mode to work. Location is treated as a binary choice between the center city and the suburbs; no model is ever actually estimated for this choice. Instead, different auto ownership and mode to work models are estimated for households conditional on their location. Variables in the auto ownership model include the number of workers, a life cycle dummy variable, a dummy variable indicating if the head of the household works outside the residence zone, a CBD workplace dummy variable, household income, transit accessibility measured relative to that of the car mode, walking time to transit, and auto insurance costs. These models were estimated using Boston home interview survey data.

c - joint models

Only one previously estimated disaggregate joint mobility choice model could be found in the literature. This model, developed by
Cambridge Systematics, Inc. (1974), focuses on the joint choice of automobile ownership and mode to work. Using the logit model and data from the 1968 Washington, D.C. home interview survey, a variety of models were estimated for various life cycle and occupation groups. These models include variables representing the following factors:

(1) level of service to work—car operating cost, in-vehicle time and out-of-vehicle time for car and transit, fare;

(2) shopping accessibility—both car and transit;

(3) auto ownership costs—$1,000 per auto;

(4) household characteristics—income, household size, number of licensed drivers, housing type, workplace of head.

The Cambridge Systematics models rely on a choice hierarchy which is directly analogous to the one in Figure 2.2 but separates auto ownership and mode to work as "medium" range decisions from long range decisions of location and housing. The work described in this thesis is a direct outgrowth of the Cambridge Systematics research and expands the range of options and the scope of the study to include the housing and location choices.

The Cambridge Systematics study demonstrated the importance of market segmentation in model formulation. Nine different models were developed for various occupation classes and family life stages. The segmentation scheme was explicitly designed to create groups with homogeneous utility functions, and the model structures varied among the market segments to reflect behavioral differences. The coefficient estimates for the models showed significant and logically consistent variations across household types. Furthermore, comparisons between the
separate models and a single model estimated from a pooled data sample indicated that substantial biases may result from ignoring heterogeneity in preferences.

**Summary of Previous Models**

Table 3.2 summarizes the behavioral models reviewed in this section, and presents the following information:

1. the name of the researcher(s);
2. the year of publication of the reference for the model;
3. the type of methodology used;
4. the model structure;
5. the mobility choices considered; and
6. the data source used for estimation.

**3.7 Summary and Conclusion of Literature Review**

This chapter examined a small but hopefully representative sample of the enormous existing body of research into the determinants of urban form and residential location. Theoretical studies, sociological studies, impact oriented models, comprehensive land use models, and behavioral models of mobility decisions were each reviewed.

All these previous research efforts, regardless of the approach used, provide some insight into the way in which households make their mobility decisions. Clearly, no one methodology dominates the others in terms of the information it has yielded. Each is part of a body of research which has its greatest value when taken as a whole.

A number of important hypotheses are supported by all of the studies.
<table>
<thead>
<tr>
<th>Researcher</th>
<th>Year</th>
<th>Methodology</th>
<th>Model Structure</th>
<th>Mobility Choices Considered</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mayo</td>
<td>1971</td>
<td>linear probability</td>
<td>unidimensional</td>
<td>location</td>
<td>Milwaukee</td>
</tr>
<tr>
<td>Wheaton</td>
<td>1974</td>
<td>regression (linear &amp; non-linear)</td>
<td>unidimensional</td>
<td>location</td>
<td>San Francisco</td>
</tr>
<tr>
<td>Quigley</td>
<td>1973</td>
<td>logit</td>
<td>unidimensional</td>
<td>housing</td>
<td>Pittsburgh</td>
</tr>
<tr>
<td>Ellis</td>
<td>1966</td>
<td>discriminant analysis</td>
<td>unidimensional</td>
<td>housing</td>
<td>Tucson</td>
</tr>
<tr>
<td>Burns et al.</td>
<td>1975</td>
<td>logit</td>
<td>unidimensional</td>
<td>auto ownership</td>
<td>Detroit</td>
</tr>
<tr>
<td>Hoxie</td>
<td>1970</td>
<td>discriminant analysis</td>
<td>unidimensional</td>
<td>auto ownership</td>
<td>Boston</td>
</tr>
<tr>
<td>Kain</td>
<td>1964</td>
<td>regression</td>
<td>sequential</td>
<td>entire mobility bundle</td>
<td>Detroit</td>
</tr>
<tr>
<td>Aldana</td>
<td>1971</td>
<td>discriminant analysis</td>
<td>sequential</td>
<td>location, auto ownership, mode</td>
<td>Boston</td>
</tr>
<tr>
<td>Cambridge Systematics</td>
<td>1974</td>
<td>logit</td>
<td>joint</td>
<td>auto ownership and mode</td>
<td>Washington, D.C.</td>
</tr>
</tbody>
</table>
First, household mobility decisions are extremely complex in nature; they are determined by a large number of factors all of which are potentially significant. Second, all of the studies view the work trip as being of major significance to the household's location decision, though its importance relative to other factors is an area of uncertainty. Finally, households with different socioeconomic characteristics appear to have different preferences. Race, life cycle, occupation, income, family size, and the number of licensed drivers may all influence the mobility choice.

In the following chapter, an attempt is made to consider some of the above effects in a consistent, albeit oversimplified, framework. Using the basic theory of Alonso briefly described in this chapter, a range of analytic and qualitative extensions is proposed and explored.
4.1 A Behavioral Analysis of Mobility Decisions

Unlike much simpler consumer behavior processes, the way in which households select among alternative mobility bundles is poorly understood. Goods such as sugar and flour are by their very nature extremely homogeneous; one pound is for most purposes a perfect substitute for another pound. Yet mobility bundles consisting of alternative location/housing/automobile ownership/mode to work combinations are quite the opposite. The usual demand equations in which some quantity measure is a function of various prices and household income are simply not realistic behavioral representations of the basic household decision process.

This chapter attempts to develop a broad behavioral framework for describing mobility decisions at the disaggregate level. This framework will provide a useful tool for exploring various possible behavioral factors and for expressing some previously developed hypotheses about mobility choices. While the framework relies on an established analytic theory, no claims for its mathematical elegance are made. The objective in developing the behavioral framework is to provide a mechanism for expressing in simple and consistent terms the working hypotheses upon which the empirical study described later in this thesis relies.
The next section presents a modified version of a portion of a theoretical location model developed by Alonso (1964) from which the analysis in the remainder of this chapter is derived. The principal change made in restructuring Alonso's model is the inclusion of a time budget and the explicit representation of the leisure/travel time tradeoff.

Using the analytic approach of this model as a starting point, various assumptions are relaxed and their implications are considered in the following sections. Throughout most of the chapter, only the choice of residential location is considered; however, the conclusions are all directly extendable to the more general case of mobility bundles.

The Alonso model and the extension proposed are based on a consumer theory approach. Thus, households choose amounts of commodities to consume rather than selecting alternative bundles as discussed in Chapter 2. In section 4.1, a choice theory generalization is presented, and the way in which the basic consumer theory implications can be conceptually extended to a choice theory approach are discussed. The analytic convenience of consumer theory is used as a tool to express hypotheses about behavior which would be extremely cumbersome in choice theory. However, because of the close interrelationship between these two methodologies, no great loss in generality results.

4.2 An Extension of The Basic Alonso Model: Residential Bid Rent Curves

The earliest theoretical work in location was performed by von Thunen (1842) in a study of how farmers make crop planting decisions when there
exists a single market center at which various crops could be sold at known prices. When transportation costs are accounted for, von Thunen demonstrated that farmers tended to plant crops in a way which maximized their individual profits. However, this theory was not extended to consider the location decisions of households until work a century later by Alonso and others demonstrated that the basic approach could usefully be applied to model the structure of urban areas.

The Alonso model begins with the assumption that there exists a single city center (i.e., the central business district, or CBD) which is the source of all employment and other non-residential economic activity. The CBD is surrounded by a flat, featureless, infinite plain on which households locate.* This plain is characterized by a transportation system in which travel time per unit of distance is equal in all directions.

Every household is assumed to have a utility function which determines its evaluation of alternative combinations of three goods: the amount of land used for housing (denoted as q), the distance to the central city (denoted as t), and a composite of all other commodities which the household consumes (denoted as z). This utility function will be denoted as follows:

\[ U = U(q, t, z). \]  

Eqn. 4.1

*This version of the Alonso model is the simplest form. Variations which consider some non-central employment and non-uniform transportation systems are discussed by Alonso, but only as an aside to the overall theory.
The household also has a fixed annual income, $Y$, which it can allocate to expenditures on land, travel, or other commodities. If this income is spent, then the following budget constraint holds:

$$Y = P(t)q + k(t) + p_z z$$

Eqn. 4.2

where:

- $P(t)$ is the price of land at distance $t$ from the city center;
- $k(t)$ is the commuting costs from $t$ to the city center;
- $p_z$ is the composite price of a unit of $z$, the composite commodity.

The units of $Y$ and the prices $P(t)$, $k(t)$ and $p_z$ are by definition consistent.

Each household is assumed to attempt to maximize its utility. Thus, the consumer allocation problem can be expressed as a constrained maximization problem, which when solved produces equilibrium values of the decision variables $q$, $t$, and $z$.

A thorny theoretical problem in this formulation is that distance, which appear in the utility function, is a proxy for commuting time. This imposes the requirement of uniform travel times over the entire city. More significantly, the model provides no causal mechanism by which the household allocates a scarce resource, its available time, to activities.

Some relatively minor modifications to the Alonso model can eliminate both these problems. The revised model requires the specification of a
different utility function as follows:

\[ U(q,f(t),z,L) \] \hspace{1cm} \text{Eqn. 4.3}

where:

- \( f(t) \) is the time needed to traverse distance \( t \) and is a monotonically increasing function of \( t \), and
- \( L \) is the leisure time available to the location decision-maker, exclusive of working and commuting time.

Furthermore, following the development of Johnson (1966), an additional constraint on the household's possible choices resulting from a limit on its total available time in the period of decision can be expressed as

\[ T = E + L + f(t) \] \hspace{1cm} \text{Eqn. 4.4}

where:

- \( T \) is the total time available, and
- \( E \) is the time spent working, which is assumed to be exogenously determined.

The household's location decision is then the solution to the following constrained utility maximization problem.

\[
\begin{align*}
\text{Max } U &= U(q,f(t),z,L) \\
\text{s.t. } Y &= P(t)q + k(t) + p_z z \\
T &= E + L + f(t)
\end{align*}
\] \hspace{1cm} \text{Eqn. 4.5}

*Johnson's model, which is primarily oriented towards explaining leisure trip behavior assuming a fixed workplace, allows working hours to be freely adjusted by the household. However, this seems somewhat unrealistic and would unnecessarily complicate this analysis.
This problem can be rewritten in terms of Lagrange multipliers as

$$\text{Max } Q = U(q,f(t),z,L) - \lambda (Y-P(t)q-k(t)-p_z z)$$
$$- \mu (T-E-L-f(t))$$  \text{ Eqn. 4.6}

The first order conditions for solving this problem are*

$$\frac{\partial Q}{\partial q} = U_q + \lambda p(t) = 0$$  \text{ Eqn. 4.7}

$$\frac{\partial Q}{\partial t} = U_t + \lambda q p'(t) + \lambda k'(t) + \mu f'(t) = 0$$  \text{ Eqn. 4.8}

$$\frac{\partial Q}{\partial z} = U_z + \lambda p_z = 0$$  \text{ Eqn. 4.9}

$$\frac{\partial Q}{\partial L} = U_L + \mu = 0$$  \text{ Eqn. 4.10}

plus the two original constraints.

Simple algebraic manipulation of these conditions gives rise to the following:

$$\frac{U_q}{U_z} = \frac{P(t)}{p_z}$$  \text{ Eqn. 4.11}

$$\frac{U_L}{U_z} = \frac{\mu}{\lambda p_z}$$  \text{ Eqn. 4.12}

$$\frac{U_t}{U_z} = \frac{\lambda q p'(t) + \lambda k'(t) + \mu f'(t)}{\lambda p_z}$$  \text{ Eqn. 4.13}

*The notation $U_x$ indicates the first partial of $U$ with respect to the variable $x$, and the prime in the functions $k'(t)$ and $f'(t)$ indicates a derivative with respect to $t$, the distance to the CBD.
In order for these first order conditions to provide intuitive insight into location decisions, some behavioral interpretation is necessary. The first condition,

$$\frac{U_q}{U_z} = \frac{P(t)}{p_z} \quad \text{Eqn. 4.14}$$

implies that at equilibrium the marginal rate of substitution between land (q) and other commodities (z) equals the ratio of their respective marginal costs. This result is identical to the condition which results from conventional consumer theory. In diagrammatic terms, the tradeoff between q and z can be represented as an indifference map for a given value of distance $t^*$ and leisure time $L^*$, with various curves in the (q,z) plane representing different levels of utility as in Figure 4.1. The line B is the budget constraint for a constant value of $k(t)$ equal to $k(t^*)^*$. The point of equilibrium, labelled A, is the point which maximizes utility and satisfies the constraints, and hence also satisfies the first order conditions.

The development of the intuitive meaning of the second and third equations requires a somewhat more complex analysis. First, by hypothesis households have an aversion to travel time, which is a monotonic function of distance, t. Hence, unlike $U_q$, $U_z$, and $U_L$, the partial derivative $U_t$ is negative. Stated more simply, all else being equal, the further from the CBD a household locates the worse off it is.

*If distance is fixed then the leisure time decision has no effect on the optimum q or z, and the total time constraint does not affect the household's choice in that plane.
Figure 4.1

CONSUMER EQUILIBRIUM FOR LAND AND COMPOSITE GOOD
Furthermore, by definition both \( k'(t) \) and \( f'(t) \) are positive. Using this information and substituting equation 4.13 into 4.14, the first order conditions imply that

\[
\frac{U_T}{U_Z} = \frac{\left[ qP'(t) + k'(t) \right] / P_Z + U_L f'(t) / U_Z}{U_Z} \tag{Eqn. 4.15}
\]

Since \( U_L / U_Z \) is greater than zero and \( U_T / U_Z \) is less than zero, the above result implies that the expression

\[
qP'(t) + k'(t) \tag{Eqn. 4.16}
\]

must be negative. Thus, if any feasible equilibrium exists

\[
qP'(t) < 0 \tag{Eqn. 4.17}
\]

and

\[
P'(t) < 0. \tag{Eqn. 4.18}
\]

This demonstrates that the slope of the price of land must be decreasing as a function of distance around the point at which the consumer is in equilibrium.

An even stronger conclusion results from the inequality that

\[
-qP'(t) > k'(t) \tag{Eqn. 4.19}
\]

at the households equilibrium point. The term \(-qP'(t)\) represents the savings due to cheaper land associated with moving one infinitesimal unit dt away from the CBD, while the term \(k'(t)\) is the corresponding cost of such a move due to a change in commuting costs. Thus, the above equation implies that a household will only settle where the marginal savings derived from cheaper land are greater than marginal costs of commuting.
While this result is in itself both theoretically appealing and provides some insights into the possible way in which models of residential location decision processes might be developed, the greatest usefulness in Alonso's original work lies in the further development of so-called "bid rent" curves. These curves relate the greatest rent a household would be willing to pay for land to the distance of that land from the central city, where along any single curve the utility is held constant.

In order to derive these curves from the revised model presented above, consider a situation in which the utility for a household is arbitrarily fixed at some level $U_0$, and the distance from the CBD is fixed at some value $t_0$. What is the greatest rent, denoted as $p_0$, the household will pay for that site? Since utility is fixed at a constant $U_0$,

$$dU_0 = 0 = U_z dz + U_q dq + U_L dL$$  \hspace{1cm} \text{Eqn. 4.20}

Further, the budget constraint becomes

$$Y = p_z Z + p_0 q + k(t_0).$$  \hspace{1cm} \text{Eqn. 4.21}

Thus, at the maximum utility point for the household

$$\frac{U_q}{U_z} = \frac{p_0}{p_z}$$  \hspace{1cm} \text{Eqn. 4.22}

which is a simple variation of the first order condition obtained in Equation 4.11. Similarly, in the time constraint, if $t$ is fixed at $t_0$
then

\[ T = E + L + f(t_0), \quad \text{Eqn. 4.23} \]

which implies that \( L \) is fixed at \( L_0 = T - E - f(t_0) \), and that consequently \( dL \) is zero. Using the definition of the utility function, the value of the utility to the household would be

\[ U_0 = U(z, q, f(t_0), L_0) \quad \text{Eqn. 4.24} \]

The four equations, 4.20, 4.21, 4.22 and 4.23 provide sufficient information to solve for the four unknown, \( q, z, L \) and \( p_0 \). It is hence possible to express \( p_0 \) as a function of \( t_0 \) and \( U_0 \). This functional relationship will be written as

\[ p_0 = p_1(t_0) \\left\{ U_0 \right\} \quad \text{Eqn. 4.25} \]

where the notation \( \left\{ U_0 \right\} \) indicates that the particular form of the relationship between \( p_0 \) and \( t_0 \) depends on the value of \( U \). Alonso then argues that instead of regarding the distance \( t_0 \) as a given constant, one can regard it as a variable \( t \), and find the bid price at any location, \( p_1(t) \). In the revised model, this gives rise to four equilibrium equations as follows*

\[ U_0 = U(z, q, f(t), L) \quad \text{Eqn. 4.26} \]
\[ Y = p_z z + p_1(t)q + k(t) \quad \text{Eqn. 4.27} \]
\[ \frac{U_q}{U_z} = \frac{p_1(t)}{p_z} \quad \text{Eqn. 4.28} \]
\[ L = T - E - f(t) \quad \text{Eqn. 4.29} \]

*These equations differ from those described previously primarily in that \( p_1(t) \) is the price a household would bid for a location, while \( P(t) \) is the prevailing market price the household faces over which it has no control.
These provide four equations for five variables: \( z, q, p_1(t), L \) and \( t \).

Thus, any four variables can be expressed in terms of the remaining one. Choosing \( t \) as the remaining variable, one can express the bid rent curve as

*\[
p_1(t) \equiv U_0. \tag{Eqn. 4.30}
\]*

Figure 4.2 depicts a possible set of general bid rent curves. Curves further from the origin correspond to lower levels of utility since, at a given distance, higher curves correspond to higher land prices. In the Alonso model, bid rent curves have the following properties, all of which apply to the above model, since Alonso's proofs are directly extendable.

1. They are single valued (i.e., at any location \( t \) there is only a single value of \( p_1(t) \) which results in utility \( U_0 \).)
2. For any household, bid rent curves never cross.
3. The slope of any bid rent curve is always negative.

For example, Alonso hypothesized that low income households have steeper bid rent curves than high income households because of differences in preferences for space and distance in their utility functions. This situation is depicted in Figure 4.3, where group 1 and group 2 have low and high incomes respectively. The resulting pattern of location under such conditions will have low income households, group 1, locating

*Alternatively, an indirect utility function, \( U'(t, p) \), could be derived as in inverse function of the bid rents. Such a utility function represents the greatest utility a household can obtain from a given location, \( t \), at a given price, \( p \). This function has a direct analogue in choice theory which will be developed in Chapter 5, and provides the basic linkage between consumer theory and choice theory which makes the analysis in this chapter relevant to the study.*
Figure 4.2

BID RENT CURVES
Figure 4.3

A FEASIBLE RENT SURFACE FOR TWO GROUPS
near the CBD presumably at fairly high density, and high income households, group 2, locating at the fringe of the city at low density. The boundary of the city is the point at which no household is willing to outbid agricultural users for land. This minimal land price, \( P_A \) in Figure 4.3, pertains to the entire city. Since each household must ultimately locate, there is one feasible price (or rent surface) which exhausts all land. This surface is indicated as a dark band along the bid rent curves in Figure 4.3.

4.3 Non-CBD Workplaces and Spatial Opportunities in the Alonso Model

While quite simple in structure, the above model provides a number of important insights into location decisions. Households can tradeoff near locations with high prices (and therefore lower land and other consumption) against living closer to the CBD. However, the basic theory has a number of unrealistic assumptions.

One such assumption is the existence of a single central business district at which all employment and other non-residential activity are assumed to be located. In modern cities, there has been a significant trend towards a dispersal of employment, particularly in the manufacturing industries (Moses and Williamson, 1967). For example, in the city used in this study, Washington, D.C., over 48.4% of all employment in 1968 was located outside the CBD. According to a New York Times (Rosenthal, 1972) analysis of the 1970 census, over half the employment in the 15 largest cities lies outside the city limits.
Thus, while monocentric models allow an analytic solution to the equilibrium land market to be found, they also result in representations of consumer behavior which eliminate important behavioral characteristics. Consider, for example, a household in which the primary worker is employed at some point $W$ other than the CBD. For such a household, possible bid rent curves along the line connecting the center with point $W$ might be curves $BR_1$ or $BR_2$ as depicted in Figure 4.4. For this household, the central business district may not be the place for which it would pay the greatest rent. Rather, if one for the moment ignores the effect of the central city as an attractor of non-work trips, the bid rent curves would be symmetric around the point $W$. Thus, $W$ would simply replace the CBD as the focal point of the household's locational preferences, and the variable $f(t)$ in the utility function would be the travel time to $W$ rather than the travel time to the CBD.

Note, however, that in Figure 4.4 neither of the possible bid rent curves, $BR_1$ and $BR_2$, is drawn symmetric around the CBD or the workplace $W$. This is because location with respect to the central business district is still likely to be an important factor in the household's location decision even when the work trip destination is outside it. Shopping, social, and recreational activities will still be clustered there, and households will, all else being equal, still prefer living closer to the CBD than further away.

*Because any bid rent curve is part of an entire family, $BR_2$ is drawn below $BR_1$ in this figure without loss of generality.
Figure 4.4

BID RENT CURVES FOR NON-CBD WORKPLACE
The shape of the bid rent curves in this diagram, like most of those which follow, is only one of a number of possibilities. While the slope must always be negative from the CBD to the left and from W to the right, between W and CBD its value is ambiguous. To demonstrate this intuitively, consider two households in which the primary worker's job location is at W, one which rarely travels to the CBD and another which travels there frequently. The former household might have a bid rent curve such as BR₂, while the latter's would be like BR₁. In analytic terms, it would be possible to include another travel time term, \( f(t_{CBD}) \), in the household's utility function which represents travel time to the CBD (or to any other point). This would give rise to another first order condition in the household's utility maximization, another travel cost in the budget constraint, and another time factor in the time constraint. The relative effect of \( t \) and \( t_{CBD} \) in the utility function would determine whether the household's bid rent curve was closer to BR₁ or BR₂.

While it is possible to display this and other location situations in terms of bids for sites not along the line connecting the CBD and job location W, little generality is gained. In the three dimensional case, the points of interest, W and the CBD, are "hills" (or at least places of reduced slope) along the bid rent surface, and locations close to either of them tend to have bid rents which are higher than points farther from both. Again, there is a basic ambiguity in the relative bid rents for points closer to one than the other, depending on the form of the household's utility function.
The effect of travel to non-work destinations, or as discussed in Chapter 2, so-called spatial opportunities, applies to places other than the CBD. For example, suppose there is a regional shopping center at point S, which for convenience will be assumed to lie along a line from W to the CBD. Two possible rent curves in this situation are displayed in Figure 4.5, where BR\textsubscript{1} characterizes CBD oriented shoppers and BR\textsubscript{2} characterizes users of the regional center.

The regional shopping center would tend to elevate the bid rent surface in its vicinity from what it would be if the center did not exist. As before, the size of the effect would be related to the nature of the household's utility function.

4.4 Travel Time and Congestion

Idealized urban economic models such as Alonso's have frequently invoked the assumption that travel speeds are uniform in all directions. However, in reality, urban areas typically are characterized by the existence of transportation corridors, i.e. directions of travel which are well served by the transportation network. In many cases, these corridors are aligned radially with respect to the central business district, and thus for CBD workers travel time to work might be usefully approximated as being uniform. However, for households with non-CBD workers, travel time may not be closely related to radial distance.

*If the workplace, the CBD and the regional shopping center were not along the same line, the bid rent curve could be depicted in three dimensions. However, this would not add substantially to the basic concept of incorporating the effects of non-work activity centers into the bid rent curves.
Figure 4.5

BID RENT CURVES FOR A REGIONAL ACTIVITY CENTER
Travel time is also composed of a number of different components. At a minimum, one can consider out-of-vehicle time (i.e., excess time) separately from in-vehicle time. Furthermore, time spent on a transit vehicle may well be different from time spent in an automobile. In order to capture the full range of effects of transportation on location in an empirical study, these components of travel time should be separated. For convenience in this theoretical analysis, however, the generic time term in the utility function will be treated as a function of all relevant time components.

Perhaps the most widely examined factor affecting travel time in cities is congestion. This phenomenon, explicitly modelled by Solow and Vickrey (1971) and Yellin (1974), is clearly not uniformly distributed over the urban area. The central business district is the focus of a much greater number of trips per unit area than any other, and is hence almost always congested at peak hours. It is therefore reasonable to assume that average travel speed decreases due to congestion as one gets closer to the CBD. Under such conditions, trips of equal distance will generally have unequal travel times and will therefore be characterized by different bid rents. This can be represented in the model developed in Section 4.2 by defining the travel time function, f(t), to have its second derivative with respect to t less than or equal to zero.
A simple possible situation is one in which the travel speed increases linearly with the distance from the CBD until some point, beyond which it is uniform. This type of congestion is depicted in Figure 4.6, along with a uniform speed which equals the congested speed at point $C_1$. Point $C_2$ on this curve corresponds to the boundary of the uncongested region of the city.

Figure 4.7 depicts possible bid rent curves for both the uniform travel time and congested situations for a household whose job location is at point W. Unlike the previous example, the bid rent curves $BR_U$ and $BR_C$ are for the same household and correspond to the same level of utility in the uncongested and congested cases respectively. $BR_U$ lies above $BR_C$ near the downtown, where congested travel speeds are lower than uncongested ones.

The joint introduction of both congestion and a non-CBD workplace creates a great number of ambiguities in the shape of the curves. For example, consider two points equidistant from the workplace W but less than distance $(W-C_2)$ from W. For these points, one can unambiguously state that the bid rent curves will be higher (for a given level of utility) on the CBD side of W than on the opposite side due to the proximity of the non-work spatial opportunities associated with locations near the CBD. However, in the case of locations greater than distance $W-C_2$ from W, households with very strong preferences for accessibility
Figure 4.6

CONGESTED TRAVEL SPEED NEAR CBD
Figure 4.7

BID RENT CURVES FOR UNIFORM AND CONGESTED TRAVEL TIMES
to the workplace might pay more for locations to the right of W, since they thereby avoid the added travel time due to congestion. Such households will have bid rent curves which are more steeply downward sloped to the CBD side of W. Conversely, households with relatively strong preferences for access to the CBD will bid more for locations on the CBD side of W.

4.5 Urban Services: The Tiebout Hypothesis

A further simplification made in the Alonso model which requires some re-examination is the assumption that a city can meaningfully be treated as a flat, featureless plain. For better or worse, American cities generally consist of a large number of distinct jurisdictions, each of which may have different taxes, schools, traffic regulations, police, ambulances, hospitals, and other attributes which distinguish it from its neighbors.* Thus, two locations which are in all ways identical except that they are in different jurisdictions may have very different bid rents.

The question of how jurisdictional attributes, which can be collectively termed the tax-service package, affect rents has been the focus of considerable theoretical and empirical research. Tiebout (1956) proposed that the existence of many jurisdictions tended to result in households selecting locations with tax-service packages which suited their preferences. Under this hypothesis, households which are "education oriented" would purposely seek out jurisdictions with high per pupil expenditures, *

---

*Issues of topography are ignored here for the sake of clarity.
and their bid rent curves would be higher in such jurisdictions. Conversely, other households (perhaps those without children) which are not education oriented might have significantly lower bid rent curves in these jurisdictions, since high school expenditures would tend to result in high tax rates. Under the Tiebout hypothesis, the marketplace would tend to operate so that households would find the locations they prefer.*

The possible effect of jurisdictional attributes on the bid rent curves is illustrated in Figure 4.8. As with the previous bid rent curves, the household is assumed to have a primary worker with job location W, which for generality is a point other than the CBD. Regions \( J_1 \), \( J_2 \) and \( J_3 \) along the horizontal axis correspond to three different jurisdictions.

Two bid rent curves, labelled \( BR_1 \) and \( BR_2 \), are depicted in Figure 4.8. In this diagram, bid rent curve \( BR_1 \) is drawn for a household with preferences for the tax-service package of jurisdiction \( J_1 \) and \( BR_2 \) is drawn for households with preferences for jurisdiction \( J_2 \). Both households are assumed to be indifferent with respect to the taxes and services in \( J_3 \). Since these curves are for households with different utility functions, \( BR_1 \) is drawn below \( BR_2 \) without loss of generality.

Note that both these bid rent curves are discontinuous. The size of the discontinuity at the jurisdiction boundaries reflects the greatest amount the household is willing to pay for the differential tax-service

*As will be considered in Section 4.7, this process may not be as universally beneficial as the above discussion would indicate, since the exclusion of some households from some opportunities may have strong negative distributional impacts.
Figure 4.8

BID RENT CURVES FOR DIFFERENT JURISDICTIONS
package. Clearly, if two jurisdictions have completely identical services but one has greater taxes, the size of the discontinuity between those jurisdictions is the tax difference, assuming taxes and rents are in equivalent units.

4.6 Pollution

Cities, as concentrated areas of economic activity, are also concentrated areas of the wastes those activities produce. Thus, urban areas are characterized by high levels of pollution. In more general economic terms, cities are areas where the incidence of uncompensated externalities is quite high.

Rothenberg (1972) distinguishes between pollution and congestion, though the two frequently have similar impacts. Congestion, as discussed previously, results from impacts that users of a system impose on one another. Pollution, however, is an impact that an activity imposes on another group for which it does not pay.

Both pollution and congestion have concerned urban economists because they result in sub-optimal resource allocation, particularly in cities where activities are highly concentrated. In intuitive terms, if in producing some product a group is imposing a cost on others for which it does not pay, it will tend to produce more than it otherwise would. This additional production is from a societal perspective inefficient for it utilizes resources which could be allocated to produce greater benefits elsewhere.
Pollution has been hypothesized to have a substantial impact on patterns of residential location. In a direct extension of the Alonso model, Stull (1971) analyzes the theoretical structure of a city in which polluters locate around the CBD and household bid rent functions include a term representing the proximity of the pollution source to their residential location.

The effect of pollution on rents has also been the subject of empirical research, Ridker and Henning (1967) constructed a model using data from St. Louis in which the level of sulphur dioxide was included as an independent variable. The results indicated that pollution may well be a significant factor in determining rents; however, since sulphur dioxide levels are a proxy for the entire range of atmospheric pollutants, some care must be used in interpreting this result.

To illustrate the effect of pollution on bid rents consider a stationary pollution source located at point P. As before, the household's bid rent curves are based on an assumed workplace W, and the points W, P and the CBD are for clarity assumed to lie on a line. A possible bid rent curve is depicted in Figure 4.9. Note that the slope of the curve near the pollution source is positive, reflecting the household's preference for living farther away from the pollution source. While this positive slope need not necessarily exist, one can state unambiguously that for points outside the plane of the diagram which are
Figure 4.9

BID RENT CURVE WITH A POLLUTION SOURCE
equidistant from W and the CBD, households will be willing to pay more for sites farther from the pollution source. Thus, the pollution source depresses the bid rent surface in its vicinity, and the magnitude of the drop depends on the relative disutility of proximity to the source.

The same logic applies to pollution sources other than the usual smoke stacks, factories, or utilities. For example, highways are sources of automobile emissions and noise, elevated transit lines result in overshadowed streets and high, though intermittent, noise levels, and the existence of prisons or other institutions which might be termed "bad neighbors" are all various types of pollution sources of some form.*

On the other hand, there may exist what could be termed "positive" pollutants, i.e. benefits which accrue to one group for which they do not pay. "Good neighbor" institutions such as some universities which supply educational and recreational services to the neighboring community or areas of vacant wooded land both may raise the bid rent surface in their proximity.

4.7 The Role of Race and Ethnicity

Race, and more generally ethnicity, has long played a major role in determining locational patterns. Historically, there has almost always been distinct ethnic groups in urban America, and the existence of these groups has greatly influenced locational decisions. However, in recent decades the effect of ethnicity has diminished; race has perhaps grown.  

*Note that in the case of transportation facilities, improved access to both employment locations and spatial opportunities will tend to increase bid rents and may either partially or totally offset pollution effects.
A great deal of research has been devoted to analyzing differences in rents and property costs that result from racial differences. Taueber and Taueber (1965) attempt to measure the degree of segregation in urban areas, and conclude that if anything, the degree to which cities are racially segregated has increased since World War II. Laurenti (1960) explores how patterns of location and race have affected property values in several urban areas. Numerous other studies have focused on other issues affected by the relationship between race and location.

The bulk of this literature rarely focuses on the underlying behavioral mechanisms which cause the observed locational patterns, rents and property values. Clearly, behavioral issues are substantially complicated by large socioeconomic differences between whites and non-whites in the United States; yet it is possible to develop some specific behavioral mechanisms which might explain observed patterns of location.

Two forms of discrimination, negative and positive, can be hypothesized to exist. Under negative discrimination, members of one group have a preference not to locate near members of another. Conversely, positive discrimination describes a situation in which group members prefer to locate near other group members, but are indifferent to the proximity of households who are not members of the same group. These two mechanisms can have substantially different behavioral implications, though both may lead to fairly similar results, i.e. segregated communities.

If each group exercises positive discrimination, it is at least in theory possible for all to increase their utility by locating in
different areas. For example, consider a room in which there are two
groups such as smokers and non-smokers, where both smokers and non-
smokers would prefer to sit near members of their own group. Both
groups are therefore better off when the room is divided.

However, this simplistic model is fairly unrealistic in the case of
racial discrimination for a number of reasons. First, whites are a
large majority in most U.S. cities. As such, they can wield greater
influence in the marketplace and locate closer to positive "pollutants"
(or externalities) and further from negative ones. This effect is rein-
forced by the differences in income distribution between blacks and
whites. Furthermore, non-competitive aspects of the locational market
can tend to reinforce segregated housing patterns and shift the cost of
those patterns to non-whites. The fragmented jurisdictional structure in
U.S. cities further reinforces the segregating forces in cities. Towns
or neighborhoods which are predominantly black tend also to be predomi-
antly poor, and therefore cannot provide the services wealthier communi-
ties can afford. As a consequence, the non-white neighborhood becomes
increasingly unattractive. Conversely, in wealthy neighborhoods high
levels of urban services can readily be provided, thereby enhancing the
desirability of the community as a location alternative. The strong
correlation between income and race therefore provides a major impetus
towards segregated locational patterns.
Negative discrimination is a still stronger force for segregated patterns. Bailey (1959) develops simple models which rely on the assumption that whites exercise negative discrimination and blacks are indifferent to the racial composition of a location. Under simple assumptions, he shows that a segregated market will result, but the costs will be borne by the white majority in terms of higher land prices. More complex models in which realtors and house owners exercise negative discrimination can result in a shifting of that cost to the group being discriminated against.

Schelling (1972) develops models which rely on less stringent assumptions about discrimination. In his model, only positive discrimination exists, and the discriminators have a distribution of tolerance described by the maximum ratio of members of the other group they will tolerate. Even under these assumptions, Schelling shows that there are major instabilities in the integrated pattern, and shifts from a fairly small range of integrated patterns will produce stable segregated ones.

Hypotheses about how the racial composition of an area will influence location choices can be expressed within the framework of bid rent curves. For example, Figure 4.10 illustrates a bid rent curve BR, for a household of one group. In this figure, there are three neighborhoods, labelled N₁, N₂ and N₃ respectively, each with a different racial composition. Neighborhood N₁ is predominantly of the same group as the household, N₂ consists predominantly of members of the other group, and
Figure 4.10

BID RENT CURVE FOR POSITIVE DISCRIMINATOR
is integrated. This curve reflects the hypothesis of positive discrimination, i.e., a preference for the proximity of the same group. For comparison, the curve for an identical but non-discriminating household is drawn in dashed lines and labelled BR₂. The effect of the CBD is for clarity ignored, and BR₂ is therefore symmetric around the job location W. In this case, the positive discriminating household would be willing to bid the same amount as the non-discriminating one in N₂, since it considers itself no worse off there. However, in N₁ and to a lesser extent in N₃, the discriminating household has an elevated bid rent, since it perceives itself as better off in the proximity of its own group members.

In Figure 4.11 the bid rent curve BR₁ is for a household exercising negative discrimination. Again, BR₂ is a curve for an identical but non-discriminating household. BR₁ coincides with BR₂ in neighborhood N₁, while it is much lower in N₂ where members of the other group predominate.

The magnitude of the discrimination effect in the integrated neighborhood, N₃, in both of the above cases depends on the degree of tolerance of the households. Some negative discriminating households may have a threshold beyond which a neighborhood is perceived as being "an other group area." In this case, if N₃ is integrated below this threshold, BR₁ and BR₂ will coincide there. Alternatively, integration may be perceived as continuous, and the size of the bid rent difference will depend on the degree of integration. Finally, it is possible that some households perceive an integrated neighborhood as better than either segregated area, and will therefore be willing to bid more for such locations.
Figure 4.11

BID RENT CURVE FOR NEGATIVE DISCRIMINATOR
In actuality, most households probably have more than one of these types of discrimination affecting their location decisions. Both negative and positive discrimination probably play an important role. Furthermore, in some cases the effect may be so strong that members of a group do not really consider certain locations as real alternatives. For example, many whites and blacks probably do not view locating in sections of the city which are for all practical purposes entirely of the opposite race as feasible options. Because it can result in extremely segregated housing patterns, racial discrimination may restrict the choices of households. When one group is a minority, this effect may be very strong. Thus, in many cities blacks face an extremely limited set of alternative locations from which they can feasibly choose. This restriction of choice will be further discussed in Chapter 6.

4.8 Households with Multiple Workers

The basic location model and all the extensions discussed to this point have assumed that each household has only a single worker whose workplace is of relevance in their location decision. However, there has been a significant trend in the United States towards multiple worker families. For example, in Washington, D.C. in 1968, nearly 30% of all households had more than one full-time worker. For such households, the single worker model may be inadequate. Furthermore, shifting social attitudes towards the participation of women in the work force will tend to make the impact of multiple workers in a household on locational and other mobility decisions greater in the future.
A number of alternative hypotheses about how households with more than one worker select their location might be proposed. The simplest hypothesis (which reduces to a single worker location decision) is to assume that there is a primary worker, or "breadwinner," whose workplace is of such importance that it completely dominates the household's choice process.* Thus, the household's location decision is assumed to be independent of the workplaces of non-primary workers.

A less restrictive hypothesis still relies on the existence of a primary worker, but secondary workers do influence residential location. Thus, the household, when making its residential mobility decision, knows the exact workplace of the primary worker and also includes the potential employment locations of any secondary workers in a composite fashion. This hypothesis can be illustrated in terms of bid rent curves by considering a household with two workers, where the breadwinner works at place $W$, and there are two potential workplaces, $W_1$ and $W_2$, for the secondary worker. A possible bid rent curve for this situation is illustrated in Figure 4.12. Workplace $W_1$ is substantially larger than $W_2$ in terms of suitable employment opportunities for the secondary worker. Thus, the bid rent curve BR is actually peaked around $W_1$, while it is still downward sloping around $W_2$. As before, the effect of the spatial opportunities associated with the CBD is ignored.

A third possible hypothesis about multiple worker households is that all workers have fixed and known places of employment which are

---

*This approach was used in the previously discussed automobile ownership models (Cambridge Systematics, Inc. 1974).
Figure 4.12

MULTIPLE WORKER HOUSEHOLDS WITH A PRIMARY WORKER
relevant in the household's mobility decisions. Thus, there are a number of travel time to work terms in the utility function, one for each worker, and corresponding travel cost and time variables in the household's budget constraints. Figure 4.13 illustrates two possible bid rent curves, labelled BR₁ and BR₂, which might result from such a decision process. The two workers are assumed to have workplaces W₁ and W₂ respectively, and the CBD is again ignored. Bid rent curve BR₁ results from the hypothesis that the first worker is still in some sense primary to the household; hence this curve is more peaked around W₁. BR₂ is drawn to reflect a complete equality in the relative importance of the two workers in the household location decision. In this case, all the time terms in the household's utility function have completely equal weight, and can therefore be summed into a total household travel time; all travel costs can similarly be summed. This greatly simplifies the form of the utility function by avoiding the need to identify each worker's travel times and costs separately.

4.9 Households with No Workers

Almost all urban location models assume that every household has workers. However, the 1968 Washington Home Interview Survey indicated that approximately 12.3% of all households had no full time workers at all. These households consist primarily of retired persons, but may also include the long term unemployed, households on welfare, or the unemployable handicapped. Such households have locational decision
Figure 4.13

BID RENT CURVES FOR TWO WORKER HOUSEHOLD:
BOTH WORKERS WITH FIXED WORKPLACE
processes which are completely unaffected by work trip level of service. Instead, the other factors such as jurisdictional attributes, access to spatial opportunities, or pollution, probably dominate their location choices.

Furthermore, for retired households in particular, the dynamics of location decisions may be very important. These households may have made location choices while one or more members was employed, and the adjustment process may be extremely slow. The location in which they may presently reside is not necessarily optimal for them; rather, it reflects the difficulties associated with relocating if one is elderly and the strong social ties which developed after living in a location for an extended period.*

Because the locational decisions of households without workers are more subtle than those of households with workers and are poorly understood, the empirical study described in the following chapters does not include any such households.

4.10 Households Without Fixed Workplaces

There are many households which have full time workers who may not have a workplace which they perceive as fixed when they select their residential location. For example, many unskilled or semi-skilled workers (constituting 8.7% of all Washington, D.C. households with workers) may not work at any one place for extended periods of time. Most

*These problems are not unique to households without workers. However, they take on added importance in such situations.
construction workers, regardless of their skill level, probably do not perceive any single site as a fixed workplace. These households will have bid rent curves which differ substantially from those for corresponding households with fixed workplaces.

Households in which workers lack a fixed workplace will perceive the proximity of potential employment in a manner similar to the one discussed in section 4.8 for some secondary workers. Thus, the proximity of potential employment as well as the size of various possible workplaces in terms of the number of relevant employment opportunities will affect bid rents. An example of two such bid rent curves in a situation in which there are two workplaces, \( W_1 \) and \( W_2 \), is depicted in Figure 4.14. Bid rent curve \( BR_1 \) is for a household where work opportunities at workplace \( W_1 \) are much greater than at \( W_2 \), while \( BR_2 \) is drawn for another household where \( W_1 \) and \( W_2 \) have equal employment. Thus, \( BR_2 \) is symmetric with respect to point \( M \), the halfway point between \( W_1 \) and \( W_2 \), while \( BR_1 \) is highly skewed towards \( W_1 \). One would anticipate that all else being equal, these bid rent curves would be much "flatter" around workplaces than those for corresponding households with a fixed workplace, since the household is never certain which workplace it will travel to.

4.11 Bid Rents and the Full Mobility Choice

The previous discussion has considered bid rents only in terms of location. However, in Chapter 2 a strong case was made for the need to consider the entire range of mobility choices including location,
Figure 4.14

BID RENTS WHEN WORKPLACE IS INDETERMINATE
housing, automobile ownership, and mode to work. In this section, an attempt is made to generalize the bid rent concept to the entire mobility bundle and to demonstrate that the continuous consumer theory approach to mobility decision modelling is less applicable than a choice theoretic perspective. To avoid an unnecessarily complex analytical discussion, the effects of pollution, congestion, jurisdiction, race, etc. will not be considered individually. However, the logic of the prior discussions of these factors can readily be extended in this development.

Suppose that in addition to location, denoted here as $l$, the household also had to select one of a number of housing types, denoted by subscript $h$, one of a number of auto ownership levels, denoted by subscript $a$, and a mode to work, denoted by subscript $m$. Attributes of each mobility bundle in each location include as before the amount of land, $q$, and the travel time from $l$ to work by mode $m$, $f(\text{lm})$, but also other attributes such as the size of the unit, spatial opportunities, and jurisdictional attributes, which will be denoted by the vector $X_{lham}$. The household's utility function can therefore be written as follows:

$$U = U(z, q, f(\text{lm}), L, X_{lham}).$$

The price of any mobility bundle $lham$ will be denoted as $p_{lham}$. As in the simple consumer theory model, the household is assumed to select the location which maximizes its utility subject to a budget constraint. Thus, the mobility problem can be written as:
\[
\begin{align*}
\text{Max} & \quad U = U(z,q,f(lm), L,x_{1\text{ham}}) \\
\text{s.t.} & \quad Y = p_z z + p_{1\text{ham}} \\
& \quad T = L + E + f(lm).
\end{align*}
\]

However, unlike the previous models discussed, it is impossible to state meaningful first order conditions for a utility maximum. Instead, the household is confronted with a problem of qualitative choice. It must select one of a mutually exclusive and collectively exhaustive set of alternatives.

The concept of bid rents, however, still has intuitive meaning. Conceptually, the household still has a maximum bid, \(p_{1\text{ham}}\), which it would make for any alternative mobility bundle. This surface is no longer continuous; instead, it is discrete. Furthermore, as in consumer theory there still exists some function of the attributes \(x_{1\text{ham}}\), the prices \(p_{1\text{ham}}\), \(f(lm)\), \(q\) and \(z\) which reflects the household's evaluation of the greatest utility it could achieve for each alternative. This function, termed an indirect utility function, will be the focus of the remainder of this study, and for simplicity all further references to utility will be assumed to apply to it rather than the actual utility.

This choice theory perspective eliminates some of the essential problems inherent in the Alonso-type model. Households rarely make their housing decisions based on the amount of land they wish to reside on. Rather, size of lot is an attribute of the mobility bundle they select. Furthermore, location, housing, auto ownership and mode to work are
actually qualitative choices. Households ultimately select one and only one option. Thus, while consumer theory, because of its analytic simplicity, can provide important behavioral insights, the underlying process is actually better described by choice theory.

4.12 Summary of Chapter 4

In this chapter, a generalized behavioral model was developed based on previous efforts by Alonso. This model, rooted in consumer theory, permitted the exploration of a range of factors which may affect household's mobility decisions, including congestion, non-CBD workplaces, spatial opportunities, pollution, tax-service packages, race and various worker types. The basic concept of bid rents was used to express the effect of each of these factors. Bid rent curves provide a consistent means of describing a variety of hypotheses about location at the dis-aggregate, or household level.

However, consumer theory has definite limitations in modelling mobility decisions, particularly when the entire range of alternatives including housing, auto ownership and mode to work are considered. These limitations arise from the fact that in reality, mobility decisions are qualitative choices. Thus, the relevant household decision process does not focus on "how much" of each commodity to consume, but "which one" to select.

The remainder of the thesis will focus on the methodological, empirical, and practical issues which must be addressed before this basic theoretical framework can be translated into a useful model, and describes the estimation results from models of mobility decisions.
Chapter 5

ISSUES IN CHOICE THEORY MODELS OF MOBILITY*

5.1 Introduction

Chapter 2 described the basic concepts of mobility and travel behavior as well as the general hierarchy of choice in which households make the entire spectrum of transportation-related decisions. In Chapter 4, some of the specific causal factors which influence location directly and housing, automobile ownership and mode to work indirectly through the joint structure of the decision process were discussed, with particular reference to how each factor might influence a household's perception of the desirability of alternative residential locations. However, both these chapters were abstract in the sense that neither considered the specific methodology which might be used in translating the overall conceptual framework of joint choice to a particular empirical model which captures the essentials of the household's basic behavioral process yet can be estimated from existing data.

This chapter presents the basic methodology used in the empirical portion of the study. This methodology, termed disaggregate choice theory, is briefly discussed in the first three sections of this chapter. The goal of this presentation is not however to exhaustively review all the analytic details of disaggregate choice models. The reader is

*Sections 5.2, 5.4 and 5.5 of this chapter have appeared in earlier versions in A Behavioral Model of Automobile Ownership and Mode of Travel, op. cit.
referred to more complete references for such discussions. Instead, this chapter focuses on a few specific and more detailed methodological questions which relate particularly to the problem of modelling mobility decisions.

Following an introduction to choice theory, three general topics are considered. First, two alternative estimation procedures, the conventional maximum likelihood and the maximum score methods are discussed and compared. Then, the issue of whether models of joint choice processes are best estimated in a single joint structure or in a sequential manner is considered. The problem of how true behavioral alternatives can meaningfully be grouped without losing the desirable properties of disaggregate models is then resolved, and two alternative grouping methods are compared.

5.2 Models of Choice Behavior*

A predictive model of consumer behavior is, in general, a mathematical representation of a decision process. As discussed in Chapter 4, traditional consumer theory is concerned with determining how much of each of a number of commodities a given consumer will select. However, some consumer decisions are better characterized by addressing the question of which of a number of available options a household will select. This perspective is the basis for choice theory models.

*Choice theories are reviewed in Luce and Suppes (1965), Charles River Associates (1972), Ben-Akiva (1973) and Manski (1973).
In the choice theory context, the behavioral unit is faced with a set of alternatives from which one is chosen. The decision-maker is also faced with a variety of constraints that determine the feasible alternatives available to him or that he can afford. This set of feasible alternatives is referred to as the choice set or the set of available alternatives. As in consumer theory, the decision-maker's selection of an alternative from the choice set is assumed to follow the principle of utility maximization. Hence, a utility value is associated with each alternative and is used to compare the alternatives; the one with the highest utility value is selected.

The general concept of qualitative choice theory can be formalized as follows:

Let i and j denote two distinct alternatives in the choice set of consumer t, i.e.

$$i, j \in A_t,$$

where $A_t$ is a finite set of mutually exclusive and collectively exhaustive alternatives such that one and only one alternative is chosen. Denote the utility of alternative i to consumer t as $U_{it}$. Alternative i will be chosen if and only if:

$$U_{it} \geq U_{jt}, \forall j \in A_t.$$

If $U_{it} = U_{jt}$ then the consumer is indifferent between alternatives i and j and the choice between i and j is indeterminate. However, since the
utility function is in most cases continuous, the probability of a "tie" is for all practical purposes zero.

In order for this modelling approach to be useful for predictive purposes, one needs to explain the relative values of the utilities in terms of actual observed quantities. Following the development of Lancaster (1966), one can assume that each alternative could be represented by a vector of attributes that completely describes it. Denote this vector for alternative \( i \) as \( Z_i \). The utility of alternative \( i \) can be written as:

\[
U_{it} = U_t(Z_i)
\]

where the function \( U_t(\; \;) \) is specific to the individual consumer \( t \). This function differs among consumers due to variation in tastes across the population. These taste differences can be explained by the socioeconomic characteristics of consumer \( t \), denoted as \( S_t \). Thus, utility of alternative \( i \) can be expressed as

\[
U_{it} = U(Z_i, S_t),
\]

where the function \( U(\; \;) \) is the utility function that applies to all alternatives and all consumers. By observing the actual choices and the choice sets of consumers, or their revealed preferences, estimates of the parameters of the utility function \( U(\; \;) \) can in theory be obtained.

As discussed in Chapter 4, the utility function, \( U(Z_i, S_t) \), can also be interpreted in the classical consumer theory sense as an indirect
utility function which is defined as the maximum level of utility that can be achieved for given income and prices. Thus, the maximum utility consumer t can derive from selecting alternative i will be a function of the "price" of alternative i, represented by the vector of attributes Z_t, and his "income", represented by S_t. The "prices" of other alternatives do not appear in this indirect utility since only one alternative is selected. The "prices" of the non-chosen alternatives do not affect the level of utility derived from the consumption of the chosen alternative. The socioeconomic characteristics, S_t, include income as required by the indirect utility interpretation as well as other characteristics that explain differences in tastes among consumers.

The above model would result in behavior which is perfectly deterministic. Every consumer would evaluate all feasible alternatives with his utility function and select that alternative which yielded the greatest level of satisfaction. However, deterministic behavior is based on the assumption that Z_t and S_t perfectly describe alternative i and consumer t, respectively. This is clearly an unrealistic assumption. There is always some uncertainty in the measurement of Z_t and S_t, and there may even be a pure element of random behavior in consumers' choices.

Manski (1973) lists four possible sources of randomness in the utility function. These are:

(1) unmeasured variations in preferences across individuals;
(2) elements of the attributes which are unobserved or omitted;
(3) measurement errors in the independent variables;
(4) the use of instrumental variables (i.e., proxy measures) for elements of Z_t or S_t.
For these reasons, two individuals confronted with alternative sets with identical values of the observed Z and S might well make different selections, since the random components of their utility function might be completely different. In other words, it is infeasible to attempt to find a utility function \( U(\cdot) \) that will correctly predict the observed choice in all cases. Since the observed utilities of necessity include random components, it is impossible to state with certainty which alternative the consumer will select; only the probability of selecting a given alternative can be determined. This formulation gives rise to a class of choice theory models termed random utility models.

In random utility models the probability of alternative \( i \) being chosen by consumer \( t \) from his choice set \( A_t \), denoted as \( P(i;A_t) \), can be written as:

\[
P(i;A_t) = \text{Prob}[U(Z_i,S_t) > U(Z_j,S_t), \forall j \in A_t]
\]

Eqn. 5.1

In words, the probability of consumer \( t \) choosing alternative \( i \) from the set of alternatives available to him is equal to the probability that the utility of alternative \( i \) to consumer \( t \) is greater than the utility of all other alternatives in his choice set.

In theory, one could select any random structure for the utility functions. However, only a limited class of such random utility functions lead to choice models which are sufficiently analytically tractable to be of any use. The most commonly used class of models are based on the
assumption of an additive disturbance term. In such models, it is
assumed the utility function can be expressed as follows:
\[ U(Z_{i_t}, S_{t}) = V(Z_{i_t}, S_{t}) + \varepsilon_{i_t} \]  
Eqn. 5.2

where \( V(Z_{i_t}, S_{t}) \) is the systematic (i.e. non-random) part of the utility
function and \( \varepsilon_{i_t} \) is an additive random component.

Substituting equation 5.2 in equation 5.1 and rearranging terms, the
choice probability of alternative \( i \) can now be written as:
\[ P(i:A_{t}) = \text{Prob}[\varepsilon_{j_t} - \varepsilon_{i_t} \leq V(Z_{i_t}, S_{t}) - V(Z_{j_t}, S_{t}) \forall j \in A_{t}] \]  
Eqn. 5.3

Equation 5.3 implies that the functional form of the choice model,
or the relationship between the variables in \( Z \) and \( S \) and the choice
probabilities, is determined by the joint distribution of the random
components. Equation 5.3 can be equivalently expressed as the following
convolution integral:
\[ P(i:A_{t}) = \int_{-\infty}^{\infty} \text{Prob}[\varepsilon_{i_t} = w, \varepsilon_{j_t} \leq w + V(Z_{i_t}, S_{t}) - V(Z_{j_t}, S_{t}), \forall j \in A_{t}] dw \]  
Eqn. 5.4

Substituting the definition
\[ w_{j_t} = w + V(Z_{i_t}, S_{t}) - V(Z_{j_t}, S_{t}) \]  
Eqn. 5.5

in (Eqn.5.4) this integral can be rewritten as
\[ P(i:A_{t}) = \int_{-\infty}^{\infty} \text{Prob}[\varepsilon_{i_t} = w_{i_t}, \varepsilon_{j_t} \leq w_{j_t}, \forall j \in A_{t}] dw \]  
Eqn. 5.6
Thus, the mathematical form of the choice model, i.e., the relationship between the systematic utility functions and the choice probabilities, is determined by the joint distribution of the random components. There is almost no a priori knowledge to suggest the form of the joint distribution of the random components inside the integral. In virtually all previous empirical applications of choice models to a multiple choice situation, it was assumed that the random elements are independently and identically distributed. Several authors have shown the restrictive nature of the independence assumption (e.g., Luce and Suppes, 1965; Tversky, 1972). In the absence of any workable choice models that do not make this independence assumption, these models can still be used but care must be taken in defining the choice set and in properly specifying the utilities. When using a choice model that makes an independence assumption, the alternatives considered in the choice set should be distinct and independent, avoiding the inclusion in the choice set of two alternatives which are almost identical or very similar in their unmeasured attributes, of two alternatives where one always dominates the other. The advantage of the assumption of independent random components is that it greatly simplifies the choice model and makes its application to practical prediction problems feasible. For this reason, the remainder of this thesis will focus solely on models which assume an independent disturbance.
5.3 The Logit Model

Within the range of possible choice models with independent additive disturbances, two functional forms, probit and logit, have been the most widely used. Of these two, only logit has been extensively used in the multiple alternative case. For this and other reasons developed later in this chapter, the logit model was selected for this study.

The logit model can be derived in a variety of ways. One possible derivation is based on the assumption that the random components of the alternatives are independently and identically distributed with the so-called Weibull distribution (also called the Type I asymptotic extreme value distribution of largest values or the Gumbel distribution).*

This distribution has the following form:

\[ P[ε \leq w] = \exp \left[ -e^{-\left(\alpha+w\right)} \right] \]  

Eqn. 5.7

The above cumulative distribution implies a density function which has as its mean at \( \gamma \) and its mode at \(-\alpha.\) * (\( \gamma \) denotes "Euler's constant and has an approximate value of 0.5777). Because the \( \varepsilon \)'s are identically distributed, the value of \( \alpha \) acts as a constant utility value and therefore does not influence the choice probabilities. For simplicity, \( \alpha \) will be assumed to be zero.

* Benjamin and Cornell (1970)
Substituting 5.7 in 5.6 results in the multinomial logit model (Charles River Associates, 1973):

\[ P(i:A_t) = \left[ \frac{e^{V(Z_i, S_t)}}{\sum_{j \in A_t} e^{V(Z_j, S_t)}} \right] \quad \text{Eqn. 5.8} \]

The coefficients of the systematic utility function, \( V() \), are estimated using observations of actual choices. Therefore, the observed dependent variable has a value of zero or one depending on whether or not the alternative was selected. In prediction, however, the model gives a set of choice probabilities between zero and one for the set of alternatives.

An assumption which makes the estimation of the model tractable, and yet is not overly restrictive, is that the function \( V() \) is linear in the parameters, as follows:

\[ V(Z_i, S_t) = \sum_{k=1}^{K} X_{itk} \cdot \Theta_k \quad \text{Eqn. 5.9} \]

and \( X_{itk} = g^k(Z_i, S_t) \). The two vectors of variables \( Z_i \) and \( S_t \) are now denoted as a single \( K \times 1 \) vector \( X_{it} \), the elements of which are finite functions, \( g^k() \), of elements in \( Z_i \) and \( S_t \). \( \Theta \) is a \( K \times 1 \) vector of parameters to be estimated.

Substituting equation 5.9 in equation 5.8 the logit model can be written as follows:
\[ P(i:A_t) = \left[ e^{X_{it} \theta} \middle/ \sum_{j \in A_t} e^{X_{jt} \theta} \right] \]  

Eqn. 5.10

The logit model in this form is one of a more general class of utility models Manski (1973) defines as LPAD, i.e., linear in the parameters with additive disturbances.

A requirement of the variables \( X_{itk} \) is that they do not take identical values for all alternatives in all choice sets. This means, for example, that income cannot be an element in \( X_{it} \) unless it is in some way transformed. This can be demonstrated in a simple example where the choice set consists of three alternatives. Suppose the utility function for any alternative \( i \) is as follows:

\[ V_{it} = bY_t + X_{it} \theta \]

where \( b \) is a constant and \( Y_t \) is the income of the household. The probability of any alternative being selected is as follows:

\[ P(i:A_t) = \frac{V_{it}}{\sum_{j \in A_t} V_{jt}} \]

\[ = \frac{bY_t + X_{it} \theta}{e^{Y_t + X_{1t} \theta} + e^{Y_t + X_{2t} \theta} + e^{Y_t + X_{3t} \theta}} \]  

Eqn. 5.11

Simple algebraic manipulation of this expression results in the following:
\[
\begin{align*}
P(i:A_t) &= \frac{bY_t}{e} X_{1t}' \Theta e^{\frac{bY_t}{e} X_{1t}' \Theta (e_{1t} + e_{2t} + e_{3t})}
\end{align*}
\]

Eqn. 5.12

Thus, if a variable which takes the same value for every alternative enters into every utility with the same coefficient, the probabilities are completely unaffected by that variable. For this reason, such variables must be defined as "alternative specific," i.e., they must apply to some utilities and not to others. For any given variable which is constant across all alternatives it is possible to define one alternative specific variable for each alternative except one, which acts as a reference point. The choice of which alternative specific variable to omit is completely arbitrary and has no influence on the forecasted choice probabilities. Note that the same line of argument also applies to constant terms in the utilities; a constant can be defined for every alternative but one.

5.4 Previous Applications of Choice Theory to Predicting Transportation Demand

The earliest application of disaggregate choice theory to a transportation demand problem which appears in the literature is a modal choice study by Warner (1962). Since then, the state of the art in disaggregate choice modelling has advanced with almost incredible speed.
In a few years, the field has progressed from simple models of two mode choice situations, to models of many modes, to models which include trip frequency, mode and destination decisions within a single joint structure. Clearly, it would be impossible to describe all of the models which contributed to this advance in the state of the art. However, it is useful to examine some of the more significant efforts in order to demonstrate the flexibility of choice theory in modelling a broad range of behavioral decisions.

The earliest disaggregate models utilized a variety of methodologies, including logit, probit, linear probability models, and discriminant analysis, to model the choice of trip-makers between auto and public transit.* Many of these studies were primarily concerned with how consumers tradeoff time spent travelling against monetary expenditures, i.e., the value of time. These models dealt primarily with work trips, and were not oriented toward the problems of utilizing the models for forecasting or planning.

A logical extension to these models was the consideration of more than two modes. Perhaps the best example of such an extension is a model which considers six distinct means of travel for work trips in the Netherlands: walk, bicycle, train, bus, auto and moped** (Richards and Ben-Akiva, 1974). Other less ambitious efforts include a three mode

** Mopeds are small motorized bicycles which require no licence to operate in Holland.
model (auto driver alone, car pool, and transit) developed by Cambridge Systematics (1974) and a three mode model (auto driver, transit and auto passenger) developed by Peat, Marwick, Mitchell and Company (1972) for Washington and San Diego work trips respectively.

Another extension of disaggregate models was the consideration of different types of choices, including trip frequency, destination choice and mode choice, within a sequential structure (CRA, 1972). In this approach, a chain of models for shopping trips was developed. The behavioral and empirical implications of using such a chain are explored in section 5.6.

Ben-Akiva (1973) showed that it is feasible to treat alternative combinations of different decisions groups such as modes and destinations within a single joint model, thereby eliminating the need for sequentially structured models. This work was later extended to include trip frequency for shopping trips within a joint model structure with much larger sample sizes (Adler and Ben-Akiva, 1975). The same concept of extending the set of relevant choices was also applied to modelling work trips in the previously cited study of mode to work and automobile ownership by Cambridge Systematics.

5.5 Maximum Likelihood vs. Maximum Score Estimation

All of the studies described in the previous section involved the estimation of choice models and all used the same estimation technique:
the maximum likelihood method. The general properties of these
estimators are discussed in Theil (1971), and will not be reiterated
here. McFadden (1968) demonstrated that for the LPAD logit model, the
maximum likelihood estimates are not only consistent, asymptotically
efficient, and asymptotically normal, but are also unique under very
general conditions. Also, the likelihood function is convex, so any
local maximum is also the global one. Furthermore, there is currently
relatively flexible and user-oriented computer software available for
finding the maximum likelihood estimates of the logit model (Cambridge
Systematics, 1974).

Recently, however, an alternative estimation procedure has been
developed by Manski (1974) which merits careful consideration. Manski's
procedure attempts to find the coefficient estimates of the utility
functions which satisfy the following conditions:*

\[
\text{Max} \quad \hat{\Theta} \quad \sum_{t=1}^{T} \sum_{i \in A_t} \delta_{it} \\
\text{Eqn. 5.13}
\]

where:

\( \hat{\Theta} \) is the vector of parameter estimates,

\( T \) is the number of observations, and

\[
\delta_{it} = \begin{cases} 
1 & \text{if } i \text{ was chosen and } v_i > v_j \forall j \\
0 & \text{otherwise}
\end{cases}
\]

*Manski extends these conditions to a more generalized case, but the
basic concept behind the estimator is the same.
This estimation technique requires a search for a set of coefficients which maximizes the number of times the greatest utility alternative was selected.

The maximum score method has a single major advantage over the more conventional maximum likelihood approach. The logit maximum likelihood estimates require for their consistency that the disturbance for every individual decision-maker and for each alternative be independently and identically distributed with the Weibull distribution. This is an extremely restrictive assumption, particularly when (as in most cases) there is no strong theoretical reason why the Weibull distribution should pertain. In contrast, Manski's maximum score estimates do not require any specific distributional assumption for consistency. Furthermore, they do not require that the disturbances be distributed identically from observation to observation. All that is necessary is that the disturbances be order-preserving, which in non-analytic terms implies that the alternative with the highest utility also has the greatest probability of being selected.

Because the maximum score estimators are consistent under a much broader range of conditions, they are far more robust than their maximum likelihood counterparts. Manski further demonstrates that logit model parameter estimates can be straightforwardly derived from the maximum score parameters if the probabilities of choice are needed for prediction purposes.
Unfortunately, in exchange for robustness the maximum score estimates give up a great deal. These estimates are neither asymptotically efficient nor normal. Thus, it is impossible to perform asymptotic statistical tests of the significance of coefficients or of linear restrictions on the parameters. Furthermore, the maximum score estimates are not unique, and there may exist many local optima.* Finally, finding maximum score estimates requires a multi-dimensional search of the parameter space, and fully tested user-oriented computer software for performing this search is not yet available.

For the above reasons, and because of the relatively successful experiences of previous researchers with the maximum likelihood estimation method, this study does not use maximum score estimation. However, further research using maximum score estimation will probably prove quite fruitful.

5.6 Recursive vs. Joint Estimation of Choice Models

A great deal of debate in the field of transportation demand analysis has resulted from the question of whether choice models involving more than one group of decisions should be estimated as a sequence of models or in a joint model (Kraft, 1974).

*Some simulation experiments by Manski indicate that local optima and multiple global optima for maximum score estimates tend to be close to one another, though this has yet to be confirmed in a large scale study.
Where there is a clear and well-defined theory of the sequence in which decisions are made, such as the block conditional structure proposed for mobility and travel choices in Chapter 2, a sequential estimation process is obviously appropriate. In this case, the joint probability \( P(MB, TP) \) of a household selecting a mobility bundle \( MB \) and a travel pattern \( TP \), is assumed to be determined by the household in a two-stage procedure. First, the household makes long-term decisions by a process modelled by \( P(MB) \), the probability of selecting each mobility choice alternative. Then, the household makes travel choices conditional on \( MB \) by a process modelled as \( P(TP | MB) \), i.e., the probability of selecting \( TP \) conditional on the selection of \( MB \).

Using both these models, the total joint probability can be derived by applying simple probability theorems. Thus,

\[
P(MB, TP) = P(MB) \cdot P(TP | MB).
\]

Within the group of mobility decisions or travel choices, however, there is no obvious ordering of choice. Despite this lack of a natural sequential structure, it is mathematically possible to express any joint probability as the product of marginals and conditionals. For example, the joint probability of a mobility bundle consisting only of location, \( l \), and housing, \( h \), might be written as follows:

\[
P(lh) = P(l) \cdot P(h | l).
\]

*For simplicity, the probabilities of choice are written without the notation for the individual consumer \( t \); however, all of the following theoretical development applies to each individual.
In this case, the utility function of a location-housing combination \( lh \) can be expressed as follows:

\[
V_{lh} = X'_{1h} \Theta_1 + X'_{1} \Theta_2 + X'_{h} \Theta_3
\]

Eqn. 5.14

where:

\( X_{1h} \) is a vector of attributes of a location-housing bundle \( lh \) which have a different value for each \( l \) and \( h \),

\( X_{1} \) is a vector of attributes which have different values for alternative locations but not housing types, and

\( X_{h} \) is a vector of attributes which have different values for alternative housing types but not locations.

In the marginal probability of location choice, \( P(l) \), the variables in \( X_{1h} \) and \( X_{h} \) which vary over different housing choices must somehow be grouped, since the housing choice is indeterminate in the location marginal probabilities. The method of grouping, termed a composition rule, has implicit within it behavioral assumptions about the sequence of choice. The most commonly used rule (CRA, 1972) is based on the expected values of variables in the conditional probabilities. Thus the composite variables could be defined as follows:

\[
\bar{X}_{h} = \sum_{h} P(h | l) X_{h}
\]

Eqn. 5.15

\[
\bar{X}_{1h} = \sum_{h} P(h | l) X_{1h}
\]
These composite variables can then be used as independent variables. CRA (1972) also used a linear function of these variables termed an inclusive, or generalized, price. This function, denoted as $GP_h$, is defined as follows:

$$GP_h = \bar{X}_h \hat{\theta}_1 + \bar{X}_{1h} \hat{\theta}_3$$

Eqn. 5.16

where $\hat{\theta}_1$ and $\hat{\theta}_3$ are estimates of $\bar{\theta}_1$ and $\bar{\theta}_3$ from the conditional model. The generalized price can be introduced as an independent variable in the marginal probabilities.

If the data setup costs were negligible (which they are generally not), the use of the sequential estimation technique could reduce the computational requirements of the model estimation. In general, it is simpler to solve a number of smaller maximization problems than one large one. However, this method of reducing the problem of estimating a joint choice model to one of estimating a series of marginals and conditionals was not used in this study for two basic reasons.

First, because composition rules have implicit in their structure behavioral assumptions about the joint choice process, the use of different sequential representations of the joint probability and different composition rules will in general lead to different estimates of the coefficients of the utility functions.* It has been shown by

---

*McFadden (1974) views the differences in coefficient estimates as resulting from differences between first order approximations of the joint utility. However, whether these differences are interpreted as resulting from a sequential behavioral assumption or different approximations, their magnitude is such so as to make a sequential estimation highly suspect.
Ben-Akiva (1973) that the behavioral implications of the estimation results can be highly dependent on the structure assumed in the estimation process.

A second reason for avoiding sequential estimation of a joint utility function is that even if no behavioral assumption were required the sequential estimates would be less efficient than the joint ones. This can be at least intuitively demonstrated by examining each stage of the sequential estimation process. Returning to the simplified example of location and housing choice, a sequential estimation would begin with the conditional probability, i.e.

\[ P(h|l) = \frac{e^{x_1h\theta_1 + x_3h\theta_3}}{\sum_{h} e^{x_1h\hat{\theta}_1 + x_3h\hat{\theta}_3}} \]  

Eqn. 5.17

The estimates obtained by maximum likelihood or maximum score methods would be consistent, since in the case of the conditional model no assumption about sequence was required to find them. However, the estimates of \( \hat{\theta}_1 \) and \( \hat{\theta}_3 \) would naturally be less efficient, since less data was used to obtain them. Thus, by using only the conditional decisions a great deal of efficiency is lost.

In the second stage of the sequential estimation process, this inefficiency is further compounded. The composite variables are created using the probabilities from the conditional choice models, and hence are functions of the conditional estimates, \( \hat{\theta}_1 \) and \( \hat{\theta}_3 \). These in turn are random variables; thus, the marginal model estimates have an error in
measurement associated with the composite variables. This error may be further increased by the use of generalized prices, which are linear functions of the random estimates, \( \hat{\Theta}_1 \) and \( \hat{\Theta}_3 \). The randomness of the parameter estimates of the conditional model, which exists even if the model is perfectly specified, results in a measurement error in later marginal choice models. This error can only tend to reduce the efficiency of the estimation process and is probably compounded when more than two states are used.

For both the above reasons, the computational advantage of the sequential estimation was deemed far less important than its disadvantages in terms of estimation efficiency and behavioral structure.

5.7 The Problem of Grouped Alternatives

The previous chapters have considered the choices of households making mobility decisions solely from a behavioral perspective. Each housing unit in an urban area was considered to be a distinct alternative, one and only one of which is selected by any given household. However, this gives rise to an unreasonable number of alternatives if one wished to estimate a model using all possible mobility bundle combinations. Even more significantly, however, is that the use of true behavioral choices at the highly disaggregate level of the individual housing unit may place severe limitations on the estimation of a truly causal model. Data about individual housing units are typically not available for any reasonable sample of locations.
In the development of choice theory it was assumed that each alternative included in the decision-maker's choice set was a true disaggregate option, i.e., for mobility decisions, a specific location, housing, auto ownership and mode to work bundle. Most readily available data, however, describe the attributes of groups of mobility bundles such as zones or tracts, and most surveys provide data only about which group a household selected rather than which specific unit. There is no a priori reason why the use of data about groups of true behavioral alternatives should produce meaningful estimates of the true household utility functions. Furthermore, for any practical forecasting application, it is unreasonable to assume that information about the literally millions of possible mobility alternatives will ever be available. Thus, there appears to be an inherent conflict between the desire to develop a truly behavioral model in which consistent estimates of households' underlying utility functions can be found and the need to use data which either currently exist or can reasonably be expected to exist in the foreseeable future.

The remainder of this chapter addresses the question of whether it is actually critical to use information about individual alternatives. The key result is that if the logit model is used, a theoretically consistent method of grouping alternative mobility choices can
be found, and that average (i.e., aggregate) values of independent variables describing these groups can be used without introducing major bias into the estimates of the utility functions.* In addition, two alternative grouping approaches, one using physical groups and the other based on attribute groups, are considered.

5.8 Requirements for Consistency of Grouped Alternatives

Consider a set of true behavioral location alternatives, $A_t$, which has members $(l_1, l_2, \ldots, l_R)$.** An example of such a set might be specific locations in an urban area. Suppose, however, that the only data which is available is information about which census tract a household actually selected.

The utility of any true behavioral choice will be denoted as

$$U_t = U_1(Z_1, S_t),$$  \hspace{1cm} \text{Eqn. 5.18}$$

where as before $t$ denotes a specific decision-maker, 1 a particular member of set $A_t$, $Z_1$ a vector of attributes describing alternative 1, and $S_t$ a vector of attributes describing the decision maker $t$.

*The term grouping is used here rather than a more descriptive term such as aggregation of alternatives in order to avoid the already widespread confusion resulting from the distinction between aggregation of variables in model estimation and aggregation of disaggregate choice behavior to make aggregate forecasts.

**Note that for simplicity the consideration of the full mobility alternative has been dropped in this theoretical development in order to reduce notational complexities. Later the issue of how the theory applies to mobility behavior will be returned to.
Now suppose the census tracts are defined to consist of groups of separate dwelling units such that each disaggregate location is in one and only one tract. Thus, the set $A_t$ can be redefined to consist of tract elements $(L_1, L_2, \ldots, L_N)$, where $R > N$. The set of all sites in tract $L_1$ will be denoted as $l \in L_1$. Assume that the utility function has the basic logit properties described in Section 5.3. In this case, the probability of household selecting alternative site $l_1$ is as follows:

$$P(l_1 : A) = e^{V_{l_1}} \sum_{l \in A} e^{V_l}$$

Eqn. 5.19

The first question to consider is, given this disaggregate probability, what is the probability of a tract being selected. This is straightforward, since the probability of a group of alternatives being selected is simply the sum of the choice probabilities over all the group members. Thus,

$$P(L_1 : A) = \sum_{l \in L_1} P(l : A) = \sum_{l \in L_1} e^{V_l} \sum_{l \in A} e^{V_l}$$

Eqn. 5.20

or if the definition

$$V_{L_1}^* = \ln \sum_{l \in L_1} e^{V_l}$$

Eqn. 5.21

is used, then

*The subscript $t$ denoting households has again been suppressed for the sake of clarity. Since all of the following analysis applies to individual households, no generality is lost.
The last equation, Eqn. 5.22, is in the logit model form. Unfortunately, in order to obtain estimates of the $V^*_L$'s, one must know the disaggregate non-random portion of the utilities for every disaggregate unit, about which data are by assumption unavailable.

The entire problem of grouped alternatives can be considered from an entirely different perspective. Returning to the original utility function at the disaggregate location alternative level, recall that

$$U_1 = U_1(Z_1, S)$$  \hspace{1cm} \text{Eqn. 5.23}

For a household selecting a location, only the attributes of the best site in the tract are relevant. It really doesn't care what the utilities of inferior alternatives are, since it only will select the best one. Thus,

$$U_{L_1} = \max_{l \in L_1} (U_1)$$  \hspace{1cm} \text{Eqn. 5.24}

This statement should not be confused with one which assumes that the utility of a group of options is equal to the maximum of the non-random components of the disaggregate utilities. Inherent in the random structure of the model is the assumption that the best alternative may sometimes have a very low value of $V(Z_1, S)$.

In summary, the combination of practical and theoretical requirements considered in this section produce four conditions for consistency which must be met if a choice model is to be developed from grouped alternatives.
(1) The sum of the probabilities within a group must equal the
group probabilities.

(2) The utility of the group must equal the maximum of
utilities of the group members.

(3) The form of the choice model for the group alternatives
must be such so as to permit computationally feasible
estimation of the parameters of the true disaggregate
utility functions.

(4) The resulting grouped utility functions must not require
information about the specific group members; only overall
summarizes such as group means should be needed.

Fortunately, a model which satisfies all four of these conditions
can be found using fairly robust assumptions. In the following
section the derivation of such a model is presented.

5.9 A Choice Model with Grouped Alternatives

Using the additive disturbance assumption of the logit form,
Equation 5.24 can be rewritten as

\[ U_l = \max_{l \in L_l} (V_l + \varepsilon_l) \]  

Eqn. 5.25

Since \( \varepsilon_l \) is distributed as Weibull, \( V_1 + \varepsilon_l \) also is; however,
the parameter of the distribution, \( \alpha \), is now \( V_1 \) rather than zero.
Furthermore, McFadden has demonstrated that the distribution of the
maximum of a set of independent Weibull distributed variables each
with parameter \( V_1 \) as also distributed as Weibull but with parameter

\[ \ln \sum_{l \in L_l} e^{V_l} \]  

Eqn. 5.26
Thus, the probability of group $L_i$ being selected can be derived in the same way as any logit model is developed, i.e.

$$P[L_i; A] = \text{Prob} \left[ U_{L_i} \geq U_{L_j} \forall j \neq i, j \in A \right] =$$

$$e^{\ln \sum_{l \in L_i} V_l} / \sum_{L_j \in A} e^{\ln \sum_{l \in L_j} V_l}$$  \text{Eqn. 5.27}

Not surprisingly, this result is identical to the one derived in the previous section by summing the choice probabilities within a group of alternatives. In fact, given the assumption that decision-makers always maximize their utility, the criteria (1) and (2) listed in Section 5.8 are merely two sides of the same coin. If the group utility is the maximum of the group member utilities, then the group probabilities must be the sum of the group member probabilities, and vice versa. Returning to the question of deriving a grouped alternative model, recall the definition

$$V_{L_i}^* = \ln \sum_{l \in L_i} V_l$$  \text{Eqn. 5.28}

and note that this is a multivariate function of the disaggregate non-stochastic utilities, which are in turn functions of the disaggregate independent variables. The function $V_{L_i}^*$ can obviously be evaluated at the group mean values of the utilities, $\overline{V}_{L_i}$. Furthermore, if the disaggregate utilities $V_l$ are linear in their parameters
\[ \bar{V}_{L_1} = \sum_{l \in L_1} P(1; L_1)V_1 = \sum_{l \in L_1} P(1; L_1) \sum_{k=1}^{K} \beta_k X_{1k} = \]

Eqn. 5.29

\[ \sum_{k=1}^{K} \beta_k \bar{X}_{L_1 k} \]

This implies that for an individual decision-maker the mean utility of the group is a linear function of the expected values of the independent variables within the groups. Furthermore, the parameters of that mean utility are the parameters of the true disaggregate utility functions.

\( \bar{V}_{L_1} \) can be evaluated at the point \( \bar{V}_{L_1} \) in at least two ways. Using the definition in Equation 5.28

\[ \bar{V}_{L_1}^{*} \mid \bar{V}_{L_1} = \ln N_1 + \bar{V}_{L_1} \]

Eqn. 5.30
where $N_1$ is the number of elements in set $L_1$. Alternatively, one can perform a Taylor expansion of the function $V^*_L$ around the point $\overline{V}_{L_1}$ which results in the following

$$V^*_{L_1} = V^*_{L_1} \frac{V_{L_1}}{\overline{V}_{L_1}} + \sum_{1 \in L_1} \frac{\delta V^*_{L_1}}{\delta V_1} \frac{\delta V_1}{\delta V_{L_1}} (V_1 - \overline{V}_{L_1}) + \text{Eqn. 5.31}$$

higher order terms.

Substituting Equations 3.30 in 5.35, one finds that

$$V^*_{L_1} = \ln N_1 + \overline{V}_{L_1} + \sum_{1 \in L_1} \left( \frac{\delta V^*_{L_1}}{\delta V_1} \right) \frac{\delta V_1}{\delta V_{L_1}} (V_1 - \overline{V}_{L_1}) + \text{Eqn. 5.32}$$

higher order terms.

To further simply this, one notes that

$$\frac{\delta V^*_{L_1}}{\delta V_1} = \frac{\delta}{\delta V_1} [\ln \sum_{1 \in L_1} e^{V_1}] \frac{\overline{V}_{L_1}}{\overline{V}_{L_1}} =$$

$$e^{V_1} \sum_{1 \in L_1} e^{V_1} \frac{\overline{V}_{L_1}}{\overline{V}_{L_1}} = p[1:L_1] \frac{\overline{V}_{L_1}}{\overline{V}_{L_1}} = 1/N_1 \quad \text{Eqn. 5.33}$$

Thus, ignoring higher order terms,
\[ V_{L_1}^* \mid V_{l_1 \in L_1} = \ln N_{l_1} + \bar{V}_{L_1} - \sum_{l_1 \in L_1} p[1:1] (\bar{V} - V_{l_1}) = \]

\[ \ln N_{l_1} + \frac{1}{N_{l_1}} \sum_{l_1 \in L_1} V_{l_1} \quad \text{Eqn. 5.35} \]

or substituting the results from Equation 5.29,

\[ V_{L_1}^* = \ln N_{l_1} + \bar{V}_{L_1} = \ln N_{l_1} + \sum_{k=1}^{K} \beta_k \bar{X}_{L_1} \quad \text{Eqn. 5.36} \]

Returning to Equation 5.27, the choice probability of the group is then

\[ p[L_1 \mid A] = e^{V_{L_1}} / \sum_{L_j \in A} e^{V_{L_j}} = \frac{e^{\ln N_{l_1} + \sum_{k=1}^{K} \beta_k \bar{X}_{L_1}}}{\sum_{L_j \in A} e^{\ln N_{l_1} + \sum_{k=1}^{K} \beta_k \bar{X}_{L_j}}} \quad \text{Eqn. 5.37} \]

In short, this lengthy mathematical exercise has resulted in the derivation of a choice model for grouped alternatives which meets all four of the criteria established in the previous section under a fairly reasonable set of assumptions. Abstracting from the mathematics, what the above proof implies is that if the logit model applies it is possible to obtain estimates of the underlying parameters of utility functions even when the only data available refers to groups of true behavioral alternatives. The resulting model is in itself a logit form, and can be estimated using only the means of the actual groups along with the size of the group, which enters the group utility function as a natural logarithm with a coefficient constrained to unity.
5.10 Assumptions of the Grouped Model

A key assumption in the derivation of the grouped alternative model was that higher order terms in the expansion of the expected maximum utility of the group could safely be ignored. A useful question to consider is whether this assumption is realistic, and under what conditions will this assumption result in unacceptably large errors in the coefficient estimates.

To evaluate the effect of the higher order terms, one can derive the difference between the true value of the group utility and that implied by the first order Taylor expansion. This is as follows:

$$\Delta = V^*_L - (\ln N + \overline{V}_L) = \ln \sum_{1 \in L} {V^1 - \ln N} - \overline{V}_L$$  
Eqn. 5.38

In order to investigate the effect of $\Delta$, consider the simple case of grouped alternatives in which there are three possible alternatives with utilities $V_1$, $V_2$ and $V_3$ respectively. Options 1 and 2 are to be grouped. $V_3$ is fixed at zero and $V_1$ and $V_2$ can be varied parametrically.

The value of $\Delta$ is zero when $V_1$ and $V_2$ are equal. This is because in such cases $V_1 = V_2 = \overline{V}$, and thus the first and higher order terms in Equation 5.36 are all zero. This result holds for grouping any number of alternatives, as long as their utilities are identical. In the two alternative case, $\Delta$ is always negative because the higher order terms in the Taylor expansion are all negative.
Figure 5.1 depicts the relationship between the absolute value of \((V_1 - V_2)\) and the absolute value of \(\Delta\). (The term \(\Delta\) is symmetric with respect to \(V_1 - V_2\)). As one would anticipate, small differences in the utilities of the alternatives being grouped result in very small errors. Under conditions where \(V_1\) and \(V_2\) differ radically, the error becomes as large as approximately 17% of the difference between \(V_1\) and \(V_2\). However, this corresponds to grouping two alternatives with choice probabilities of about .94 and .06 respectively. Thus, even in what might be termed the worst possible case, the first order Taylor expansion approximation results in errors which, while substantial, are not so large as to make the results meaningless. In more typical cases where the grouped alternatives have choice probabilities which lie between .4 and .6, the error term is less than 6% of the difference between the utilities.

Instead of considering the errors in the utilities, one could also examine the errors in the probability resulting from the grouping approximation. In this case, the error behaves somewhat differently. Since the probabilities are monotonic transformations of the utilities, in the two group case the probability error is always negative. However, the non-linearity of the transformation results in a probability error which does not behave either monotonically or symmetrically over the range of the difference between \(V_1\) and \(V_2\). Furthermore, the error is a function of the utilities of the alternative not grouped.
Figure 5.1

ERROR DUE TO TRUNCATION OF TAYLOR EXPANSION
Figure 5.2 depicts the absolute value of the probability error for selected values of $V_1$ over the range of $V_2$ considered. The worst case found (using increments in $V$ of .25) was an absolute probability error of .1079. In this case, the probabilities of the true disaggregate alternatives were .0347 and .5427 respectively, while the grouped probability using the first order Taylor expansion was .6852. Thus, the likelihood of one of the alternatives in the worst case group was about 16 times that of the other.

5.11 Physical vs. Attribute Grouping

The above discussion of grouping location alternatives considered as an example groups of sites based on their physical proximity. However, it is also possible to define groups along the dimensions of their measurable attributes. For example, rather than use tracts as the unit of grouped locations, one might consider levels of neighborhood quality such as a three level status classification of upper, middle and lower class. Other dimensions along which groups of alternatives could be defined include the distance to work, prices, or the level of urban services provided.

The attribute-based grouping approach is adopted by Ansah (1974) to define alternative destinations in a unidimensional destination choice model. In an empirical case study, he defines
Figure 5.2

PROBABILITY ERROR DUE TO TRUNCATION
the destination choice set in terms of discrete levels of distance (or travel time) and retail floor space. However, Ansah fails to recognize the need to adjust his models for the number of members in the attribute groups. As a consequence, he concludes that by developing such a choice set the difficulties of identifying the true set of relevant alternatives for each individual are avoided, since each individual has every alternative available.* If the appropriate measure of the size of the group were introduced in his models, the individual choice sets would still have to be identified and the critical advantage claimed for his approach would be lost. Ansah also attempts to incorporate all measurable attributes into the definition of the choice set, thereby creating models which consist solely of alternative specific constant terms. He relies on market segmentation to eliminate the need to include socioeconomic variables in the utility function.

In the case of mobility choice models, Ansah's approach would be somewhat awkward, since the number of relevant attributes is quite large. Furthermore, this method would require that arbitrary distinctions be drawn along continuous attributes. This may result in an important loss of information in the estimation process. However, some type of attribute grouping that did not necessarily include all relevant dimensions of the attribute space might be considered.

*This also is not necessarily true in more general choice situations even when the size of the group is ignored. For example, some types of residential areas may, through racial discrimination in the housing market, exclude minority households.
There are three possible reasons why groupings defined along attributes might be useful. First, it is possible that some computational advantages can be gained through a reduction in the number of alternatives or the number of parameters to be estimated. Second, it is conceivable that at least some of the difficulties of making aggregate forecasts from non-linear disaggregate models such as the multinomial logit could be sidestepped, since attributes would not vary a great deal within any single group.*

The final possible advantage of an attribute-based grouping is that it may actually be a behavioral approach; households may actually perceive attribute groupings in their decision-making process. Location decisions might be hypothesized to be the result of an hierarchical choice process in which the household first selects the general category of attributes it desires and then proceeds to choose a particular mobility bundle from the set which has those attributes. In this case, the choice process could be modelled at the behavioral level of attributes without requiring the grouped model at all.

The last of the above reasons, if it were true, would probably be the only compelling justification for grouping along attributes. However, it was ultimately decided that the possible gains which might result from the use of attribute groups were more than offset

*Homogeneity of attributes within a choice group greatly reduces the problems of making aggregate forecasts. (Koppelman, 1975)
by the uncertainties inherent in determining precisely what attribute groups households actually perceive in their location choice, if they perceive any at all, and the information lost in creating discrete attribute groups. For this reason, physical groups at the tract level were chosen for use in the empirical study.

5.12 Grouping Mobility Alternatives

The simple numerical examples described in section 5.10 tend to support the hypothesis that a first order Taylor series approximation of the utility of grouped alternatives is fairly accurate, particularly when the alternatives within the group are relatively homogeneous. However, careful consideration should be given to whether homogeneity of alternatives is reasonable in the case of actual mobility alternatives grouped along physical dimensions.

The empirical study in this thesis uses a number of general dimensions of grouped alternatives. Tracts were chosen for physical groups of locations primarily because of their small size and the large amount of information available on the tract level. Each census tract is designed to include approximately 4000 residents, though there exists considerable variation in any one city. At most urban densities, this results in a fairly small area. Furthermore, since census tracts are drawn to be compact, the distribution of level of service characteristics to any workplace within a tract should be acceptably small. Tract boundaries never cross jurisdictional boundaries so variations in tax-service packages and
neighborhood characteristics are minimal. Variation in auto ownership costs such as fuel prices, maintenance costs, and insurance costs should be for all practical purposes negligible within a tract. Finally, as with any small, compact, contiguous geographical area in a city, tracts will tend to have housing which is uniform in difficult to measure attributes such as prestige, architectural style, or overall housing quality resulting from maintenance and repairs.

Note that no claim is made that census tracts are in any sense an optimal grouping method. In many ways, tract boundaries are fairly arbitrary and frequently are not extensively readjusted as the city changes. Furthermore, no claim is made that households actually perceive the tract as a behavioral unit; perception of the grouping method is neither a necessary nor a sufficient condition for obtaining consistent estimates of the parameters of the utility function.

In reality, given the need for some grouping of alternatives imposed by limitations in the available data, it is unlikely that any single grouping scheme will be optimal in all dimensions. Some geographical groupings would be extremely homogeneous with respect to transportation level of service, but may cross jurisdictional boundaries. Others might be extremely homogeneous with respect to neighborhood characteristics, but would be highly concave and therefore be subject to extreme variability in transportation services. Thus, tracts, while not optimal, probably represent a fair compromise between extremes.
A second dimension in grouping alternatives is in housing type. As discussed in the theoretical development in Chapter 2, housing is a highly heterogenous commodity which can be differentiated by tenure, quality, structure type, structure age, lot size and a number of other significant characteristics. The limitations on existing disaggregate data impose severe restrictions on how those distinct alternatives must be aggregated. Thus, it was necessary to restrict the housing type choice to two tenure types (owned and rented) and three structure types (single family, garden apartments or walk-ups and high rise multi-family dwellings). The use of this grouping of alternatives may raise significant questions about within-group homogeneity. For example, owner-occupied single family dwellings vary considerably in price, quality, size, lot type and structure type.

The validity of characterizing this broad spectrum of alternatives by a single price would be extremely dubious. However, one must recall that the range of housing alternatives is only being grouped within a census tract. Thus, a different price or rent pertains to every housing type in every census tract. For this reason, the key issue is not how different a given set of housing units in the same housing type group are, but how different units are within any particular tract.
Due to factors on both the supply and demand side of the housing market, housing in any one area tends to be fairly homogeneous. Developers (presumably due to some economies of scale) tend to construct housing developments within which variation in housing attributes is minimal. It is also rare for two contiguous areas to be developed at very different times. Instead, one development tends to lead to others of similar style adjacent to it. Furthermore, land owners and developers can generally be assumed to construct the type of dwelling units which they perceive as maximizing their return. Thus, to the extent that information is equally available to all and to the extent that perceptions do not vary widely among various actors, one would anticipate that any one geographical area would be developed relatively uniformly.

All of these factors will contribute towards spatial homogeneity of housing. Clearly, some tracts will be extremely heterogeneous, perhaps to an extent which violates the assumption inherent in the first order Taylor expansion to an unacceptable degree. This may be particularly true for a tract drawn to include the classical "right side" and "wrong side" of the tracks. In such a case, the average price and other characteristics will tend to exhibit a bimodal distribution and the mean will not be a good estimate of the attributes of either type of unit. Nevertheless, in general the mean attributes of the housing types used within any tract should not introduce an unacceptable bias into the model.
Automobile ownership is also a highly grouped alternative. Autos differ by type and age, yet the model distinguishes only among alternative levels of auto ownership. Many of the attributes including cost vary widely within any set of grouped auto ownership levels. Furthermore, for practical reasons, a single alternative including all multiple auto ownership categories was defined. Unlike the case of location and housing types, no compelling reasons for expecting various auto types to be homogeneous can be presented. However, at least along some dimensions, some homogeneity seems reasonable. For example, within reasonable limits, all autos offer virtually identical travel times. Nevertheless, the auto ownership aspects of this study will probably not adequately reflect the broad variation in the type of autos which will be owned.

A final dimension of grouping alternatives is in modal choice options. Only two alternatives are considered in the model, transit and car. However, each of these actually is a grouping of more disaggregate alternatives. The auto driver car mode consists of various alternatives such as driving alone, paid car pools, long standing non-paid shared rides, intra-household shared rides, and short term arrangements resulting from temporary employment or needed auto maintenance. In general, transit can be broadly defined to include fixed route bus, commuter railroads, rapid rail and an
entire spectrum of more flexible paratransit services. In these cases, related means of travel are being grouped together and average attributes are being used. However, these modes have closely related attributes, particularly in terms of travel time. Furthermore, many of the unmeasured attributes such as overall comfort and convenience probably have fairly low within-group variability, particularly when compared with the between-group variation. Thus, this grouping of alternatives should probably produce acceptable coefficient estimates.

5.4.3 Summary of Chapter 5

This chapter has addressed a number of important issues in the application of choice theory models to household mobility decisions. The basic elements of choice theory were developed with particular reference to the multinomial logit model, which is used in the empirical study described in the following chapters.

Two techniques for estimating choice models, the maximum score and the maximum likelihood methods, were discussed and compared. The former method offers some advantages; in particular, it is quite robust with respect to the distributional assumptions it makes about the disturbance terms of the utility functions. However, the maximum score method is computationally burdensome and has unknown statistical properties. For this reason, the more conventional maximum likelihood method was selected for this study.

*Actually, at the time of the household interview survey transit in Washington consisted almost exclusively of fixed route bus.
A second issue considered was the use of sequential methods of estimating joint choice models. The sequential approach, adopted by CRA(1972), is both behaviorally unsound and produces estimates which are considerably less efficient than those resulting from a joint model estimation. For these reasons, the joint model approach was selected for this study.

The final question this chapter addressed was the problems inherent in attempting to model a behavioral process when the available data provide information about groups of relevant alternatives rather than the actual disaggregate alternatives. A set of four conditions were established in order for the results to be both empirically useful and consistent with the underlying theory of disaggregate choice behavior. Using the multinomial logit model, it was demonstrated that under relatively robust assumptions using the average attributes of the groups of alternatives in the utility functions along with an additive element which represents the natural logarithm of the number of members in the group results in a model which meets the four criteria. Furthermore, if the assumptions hold, the estimates of the grouped alternative utility function parameters are also consistent estimates of the true disaggregate utility function parameters.

This last result has important implications for a broad range of modelling situations beyond the mobility decisions considered in this thesis. In fact, almost all previously estimated models have
to some extent relied on a grouping of alternatives without ever addressing the basic theoretical issues inherent in such grouping. Destination choice models have been estimated using shopping centers or zones as the relevant destination alternative for shopping trips when in fact it is possible that specific stores may be the true disaggregate alternative (Ben-Akiva, 1973; Buro Goudappel en Coffeng and Cambridge Systematics, Inc., 1974). Similarly, joint auto ownership-mode to work models have used auto ownership levels rather than type, age or quality as the disaggregate alternatives (Cambridge Systematics, Inc.). Thus, the theoretical resolution of this problem not only provides a needed underpinning for the empirical study in this thesis, but also provides basic analytic support for a large body of prior empirical studies using disaggregate choice models.
6.1 Introduction

Any empirical study inevitably must resolve the basic dilemma that available data does not completely meet the needs of the research being undertaken. However, this problem is almost always compounded in the estimation of disaggregate choice models, since existing data sets are generally designed for aggregate analysis. Thus, in order to create a useful data base, information from a number of sources must be processed, checked and organized in a form suitable for use in the model estimation process.

The assumptions made in the course of building a data base must inevitably raise some significant questions about possible errors in the empirical study. Each assumption introduces some randomness and perhaps bias into the measurement of certain variables. The basic issue which must be addressed is whether these sources of error either individually or collectively are so large as to invalidate the conclusions drawn from the study.

Unfortunately, no one simple answer to this question exists. The magnitude of potential errors is almost impossible to evaluate without actually collecting data better than the original source used; such a task was impossible within the limited resources of this study. Furthermore, the way in which measurement errors affect the estimation of highly non-linear probabilistic choice models such as the multinomial logit is
not well understood. In the final analysis, the importance and validity of assumptions made in the data processing must remain a judgemental issue.

The problem for the researcher, therefore, is almost always one of allocating scarce resources to the task of data collection and processing. New data can be gathered or existing data can be checked. In this study, all of the data used is taken from previously gathered sources. Heavy emphasis was therefore placed on insuring the validity of the data and on eliminating invalid or questionable observations wherever possible.

This chapter considers some of the most significant issues resolved in the course of developing an appropriate data set. The problem of whether engineering estimates or perceived attributes of alternatives in the choice set is considered in the following section. Then, the data sources used to create the data base are described, and the assumptions used to resolve some potential inconsistencies among the sources are presented. The method used in subsampling and verifying observations is also discussed. Finally, a statistical profile of the most significant attributes of the data set and the possible implications of these attributes for model development are presented.

6.2 Perceived vs. Measured Attributes

The data set used in the estimation of a disaggregate model can be logically divided into three categories. First, it is necessary to know the actual choices made by a sample of decision-makers. Second, one
generally needs to know something about the decision-makers in terms of their socioeconomic characteristics and the alternatives they had available. Finally, one needs to know the attributes of the alternatives they faced when making their decisions.

Ideally, all of the data items used would be at the disaggregate level. However, as in most other studies, no such data set suitable for the estimation of mobility choice models could be found. An alternative data source is so-called "quasi-disaggregate" data, i.e. data which has been aggregated into a large number of tabulation units such that most of the variation at the disaggregate level is retained. In this study, most of the variables in the third category, attributes of the alternatives, are quasi-disaggregate.

All of the attributes used are either engineering estimates or observed values as opposed to the attributes actually perceived by the decision-maker. The debate over whether "actual" or "perceived" data should be used in the estimation of disaggregate demand models is a long-standing one. Many planners, particularly those from mathematical psychology backgrounds, have argued that the development of choice theory is entirely predicated on the assumption that individual decision-makers select the alternative which they feel will maximize their utility and that only this perceived utility should be used in the development of choice models.

This research, which relies on the use of engineering estimates of level of service and other data, is based on a profound disagreement with
this line of argument. This disagreement is predicated on two groups of reasons. First, the use of perceived data creates a number of practical problems if a model is ever to be used as a forecasting tool. More significantly, however, is that a theoretically consistent model which relies solely on engineering estimates can be developed and the question of the relationship between perceptions and actual values of variables can be implicitly incorporated into behavioral choice models.

The first practical problem with the use of perceived data is that one never knows exactly what individuals really perceived when they made their choices. All one can actually hope to know is what they reported to an interviewer. Psychologists have long recognized that such reports can be highly biased. The mechanism of so-called "cognitive dissonance" tends to result in individuals reporting perceptions so as to justify their own decisions and thereby reduce the amount of conflict they feel about their choice. (Hilgard and Atkinson, 1967).

For the transportation analyst attempting to apply a model, the use of perceived data creates still further problems. Forecasting perceptions is at best a highly dubious task. Studies by McLynn and Goodman (1973) and others have attempted to correlate reported travel times and engineering estimates for the chosen mode of travel to work in a sample of households in the Shirley Highway corridor. They concluded that there is a fairly high correlation (.66 for auto users and .70 for bus users), but ultimately decided that reported travel times were unreliable.
One can reasonably assume that the reliability of reported times for the non-chosen alternatives would have been even lower. Even if this were not the case, the proper application of a model based on reported data would actually require two distinct models. The first model would be the behavioral choice model, which relates the decision, denoted as \( D \), to the perceived characteristics of the alternatives, denoted as \( P \). This function could be written as

\[ D = f_1(P). \quad \text{Eqn. 6.1} \]

However, for use in forecasting, a second model relating perceptions to actual values of variables, denoted as \( E \), would be required. Such a model, denoted as

\[ P = f_2(E), \quad \text{Eqn. 6.2} \]

might be as complex as the original choice model.* In fact, much research, primarily by psychologists, has been devoted to examining this problem, resulting in numerous models dealing with both threshold levels and the perceived response to various levels of stimulus (Galanter, 1962).

A further problem in the use of reported attributes is that one must ask people not only about the characteristics of the alternative which they selected, but also for the characteristics of alternatives which they report as being available. In the case of simple binary choice models such as those discussed by Watson (1971), this is fairly straight-

*In a more complex dynamic model, one could consider the learning process by introducing past decisions \( D_{t-1} \) into the perception function.
forward. However, when one wishes to estimate more sophisticated models such as those being considered here, the number of alternatives about which people must report their perceptions is quite large. It is therefore not surprising that all previous disaggregate model estimation in the transportation field which has relied on reported level of service data has been restricted to simple models of mode choice; all more comprehensive models (e.g., Ben-Akiva, 1973; Adler and Ben-Akiva, 1975) have utilized engineering estimates based on conventional methods.

How, then, can one actually incorporate perceptions into a model of choice behavior which relies on engineering estimates? Suppose equation 6.2 is substituted into 6.1. The resulting model is simply

\[ D = f_1(f_2(E)) \]

Eqn. 6.3

Thus, if one considers implicitly how perceptions relate to actual values of independent variables, it is completely consistent to use the observed data rather than engineering estimates. For example, one of the earliest psychophysical theories in psychology, Fechner's Law, is that perceptions are logarithmically related to actual stimuli. In the general notation developed above, this could be denoted as

\[ P = k \cdot \ln E, \]

Eqn. 6.4

where \( k \) is a constant scaling factor. It is not surprising that many variables in models of travel behavior have appeared as logarithmic (and other) transformations of level of service attributes when engineering
estimates of the level of service are used. In creating these transformations, the role of perceptions in determining behavior is implicitly being recognized.

All of the above reasons provide a strong case for utilizing observed values of dependent variables. In developing practical analysis tools, the transportation analyst does not need to estimate models which relate perceptions to actual values; such models are of little use in design and are of only marginal value for research. Instead, it is far more reasonable to deal with perceptions implicitly in the specification of the functional form of the choice model and rely on observed and engineering values for measurement of level of service data.

6.3 Data Sources

The principal source of information about observed decisions and the characteristics of the decision-makers in the field of urban transportation is the metropolitan home interview survey. A number of possible alternative survey sources were rejected because of insufficient sample size or missing data items. The urban area ultimately selected was Washington, D.C., primarily because of the quality of the level of service data and the large base of successful prior disaggregate modelling efforts which could be drawn on in specifying mobility models. Furthermore, in the course of prior studies, the Washington D.C. survey has been extensively processed to allow for easy manipulation of the data.

* A processed file of household surveys developed by Richard L. Albright was provided for this study by Cambridge Systematics, Inc.
The Washington home interview survey (HIS) was conducted in 1968 by the Metropolitan Washington Council of Government (WCOG). 26,544 household records including information about 75,001 persons taking 179,441 trips were coded. Included in this data were the following:

1. number of cars owned
2. mode used for each trip
3. residential location (zone, district and tract)*
4. structure type
5. the work location of each person (if any)
6. type of living quarters (permanent or seasonal)
7. dwelling ownership (own or rent)
8. number of persons in the household
9. number of persons over 5 years of age in the household
10. household income
11. education level of the head of the household
12. number of licensed drivers in the household
13. the family relationship of each person to the head of the household
14. the age of each person
15. the status of each person with regard to the possession of a driver's license
16. the occupational status** of each person
17. occupation (11 categories) of each person
18. trip purpose at both origin and destination
19. time of day at both origin and destination.

* The study area for the Metropolitan Washington region was divided into 1065 zones by WCOG, which were grouped into 134 districts.

** Worker (full-time/part-time), student, housewife, retired or unemployed.
This data set, while relatively complete with respect to the choices made and household characteristics, has a number of limitations. As discussed in Chapter 5, the description of the housing choice is limited to the structure type and the type of tenure. More significantly, there is no information about the attributes of the chosen mobility bundle or of alternative bundles. This data was obtained from a range of quasi-disaggregate sources.

Transportation level of service data was taken from zonal level peak hour skim trees for highway in-vehicle time, highway distance, transit in-vehicle time, and transit out-of-vehicle time. These minimum path times and distances are based on networks coded by WCOG for the survey year. WCOG also provided transit costs derived from existing fare schedules and car operating costs from mileage based cost formulas. Highway terminal times and peak hour parking costs for the 1065 cordon area zones were also developed by WCOG.

As discussed in the previous chapter, the basic unit which will be used for the locational alternative is the census tract. Unfortunately, the zone system used by WCOG does not coincide with census tract boundaries (National Capital Region Transportation Planning Board, 1968). Tracts are substantially larger than zones since there are 605 tracts within the study area cordon as compared with 1065 zones.* However, tracts and

*Note that while there are 621 tracts in the Washington SMSA, 16 were omitted because they were outside the WCOG cordon area.
zones frequently share the same boundaries; both do not include more than one jurisdiction and both often use major transportation arteries as boundaries.

In order to obtain the transportation level of service from the interzonal matrices, each tract was matched with a representative zone whose centroid was nearest to the tract centroid. This was done by careful examination of detailed tract and zone maps rather than any particular analytic process. In cases where the geographical representative zone appeared to be vacant, another zone within the tract which was developed was selected. The representative zones were then used to determine the level of service from any work zone to any tract by using the representative interzonal service.

The above procedure is somewhat less accurate than an alternative procedure which uses a weighted function of interzonal times rather than a representative zone system. However, a weighting scheme would require vastly more computation time, and given the error already inherent in the original coding of the networks the weighted average would probably have been only marginally superior to the use of representative zones.

Data about the attributes of alternative locations and housing units were derived from the U.S. Census Fourth Count Housing A data tapes. These tape files (one each for the District of Columbia, Maryland and Virginia) provide detailed counts of a broad range of socioeconomic and housing characteristics of census tracts, including tabulations by race,
structure type, household size, etc. Most counts are from a 15% census sample, and therefore are far more reliable than data taken from tabulations of the home interview survey.

The most important data items taken from the census were rents by structure type, housing values, and demographic data such as the fraction non-white and average income of tracts. All tract data which involved monetary values was converted to 1968 dollars using appropriate deflation factors (U.S. Department of Labor, 1973). National figures were used in these computations because certain information for the Washington SMSA was not readily available. However, from the existing data it appears that the national rates of change are probably comparable to those for Washington. For example, the national consumer price index (CPI) for all costs rose 11.6% between 1968 and 1970, while the Washington CPI rose 12.3%.

It is interesting to note that while the CPI of all housing expenditures rose 14.1% over the two year period, this rise was not uniformly distributed over the entire housing sector. Rents increased only 7.5%, while home ownership costs rose by 21.6%. In the same period, average income rose by 15.2%.

The definition of structure type in the census data was somewhat different from that used in the home interview survey. In processing the census data, all structures with from 3 to 9 housing units were classified as garden-style or walk-up apartments; all larger structures were defined as high rise buildings. In contrast, the home interview survey manual defines a garden-type structure as a multi-family dwelling consisting of
either two or more multi-unit structures with a common lawn or landscaped area, and defines a walk-up as a multi-family structure of three or more stories without an elevator (Metropolitan Washington Council of Governments, 1968). High rise dwellings are defined as multi-family structures with elevators.

Another potential inconsistency is in the definition of race. The HIS defines a household as either white or non-white based on the observations of the interviewer. In processing the census, counts for non-whites were defined as the sum of all blacks and Spanish-speaking households. Whites were defined as all other households, unless separate counts for whites were available.

A third potential problem with the use of census data is that many table entries are suppressed due to the possibility of someone being able to identify an individual respondent. This occurs most frequently in multi-dimensional tabulations. These suppression codes were treated as zeros since they represent very small true entries. In the model estimation, alternatives which had undefined attributes due to this suppression were eliminated from each observation's choice set.

By using the above assumptions, it is quite possible that some high rise structures might be incorrectly classified as garden-style or walk-up apartments, or that some Oriental residents may be counted as white. Furthermore, the distribution of some attributes tends to be truncated at the extreme values, since these ranges are most frequently suppressed. However, it is unlikely that these errors are large enough to be of significance.
Other data items needed in the course of the study were gathered from a range of published sources and conversations with public officials (Welsh, 1969). Included in this category were the following:

(1) per pupil school expenditures;
(2) real estate taxes;
(3) personal property taxes;
(4) state and federal income taxes;
(5) auto insurance costs;
(6) license tag costs.

6.4 Definition of Subsample

The home interview survey was deemed too large to be used in its entirety. It was therefore decided that a subsample of the data set should be created. The best way to insure a representative subsample would be to use a random number stream with appropriate selection criteria designed to yield a certain sample size. However, the resulting observations might have workers in every possible work zone; level of service data for every zone/tract combination, of which there are nearly 645,000, would have to be maintained for the entire study.

In order to avoid this, a cluster sampling technique was used. Each workzone was treated as a cluster of observations. Fifty-eight workzones were selected at random, with selection probabilities appropriately adjusted for zone size. All households with a primary worker employed at one of these 58 zones were then selected as part of the subsample.
This subsample was then verified to insure that all data items which might be required in the model estimation were available. These checks insured that:

1. the race code was valid;
2. the income code was valid;
3. the occupation code was valid;
4. the primary worker's age code was valid;
5. the primary worker's education code was valid;
6. the household's structure type was a permanent dwelling unit other than an institution or rooming house;
7. the tenure code was valid;
8. the census tract was valid;
9. the home tract was within the cordon area;
10. the primary worker selected a mode to work of transit or automobile.

Note that the structure type criteria, number 6, limits the data set to households which have actually made long term mobility decisions rather than including residents of trailer parks, who might be located quite temporarily, or inmates of institutions, who may have had no alternatives available to them.

These ten conditions collectively eliminated 557 of 5440 potential observations. Since the criteria were imposed sequentially, it is impossible to determine whether some of these 557 eliminated observations violated more than one condition. However, at least 355 had primary workers selecting a mode other than transit or auto.
From this cluster sample a one in ten subsample was then taken. The resulting data set included 489 households with primary workers employed in 54 different work zones.

6.5 Profile of the Washington, D.C. Data Set*

Before beginning the model estimation, a number of cross-tabulations were created from the entire home interview survey and the cluster sample. Aside from providing a useful way to further check the data for validity, these cross-tabulations often yield some insight into which choices are most important to explicitly model and which seem so unlikely that they can be safely ignored or grouped with other categories.

Two types of cross-tabulations were examined. The first group focused on the demographic characteristics of the Washington survey sample with particular emphasis on occupation, number of workers, income and household size, all of which could be hypothesized to be related to a household's mobility decision. The second group of tabulations describe the outcome of the mobility decision process in terms of the housing, automobile ownership and mode to work distributions.

Table 6.1 presents the distribution of the occupation group of the primary worker by race of the household for all households in the home

*Some of the tables in this section were described in Cambridge Systematics, *op. cit.*
<table>
<thead>
<tr>
<th>Primary Worker's Occupation Code</th>
<th>Description</th>
<th>% of whites</th>
<th>% of non-whites</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>unskilled and household services</td>
<td>.5</td>
<td>5.1</td>
<td>1.4</td>
</tr>
<tr>
<td>1</td>
<td>semi-skilled</td>
<td>6.8</td>
<td>10.3</td>
<td>7.4</td>
</tr>
<tr>
<td>2</td>
<td>services, except household services</td>
<td>3.6</td>
<td>17.0</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>clerical</td>
<td>13.9</td>
<td>25.5</td>
<td>16.2</td>
</tr>
<tr>
<td>4</td>
<td>sales workers</td>
<td>5.6</td>
<td>2.4</td>
<td>5.0</td>
</tr>
<tr>
<td>5</td>
<td>craftsmen, skilled workers</td>
<td>11.0</td>
<td>12.6</td>
<td>11.3</td>
</tr>
<tr>
<td>6</td>
<td>classified</td>
<td>1.0</td>
<td>.4</td>
<td>.9</td>
</tr>
<tr>
<td>7</td>
<td>managers</td>
<td>12.2</td>
<td>6.2</td>
<td>11.0</td>
</tr>
<tr>
<td>8</td>
<td>technicians</td>
<td>5.0</td>
<td>3.7</td>
<td>4.9</td>
</tr>
<tr>
<td>9</td>
<td>professional</td>
<td>40.4</td>
<td>16.8</td>
<td>35.9</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Table 6.1

DISTRIBUTION OF PRIMARY WORKER'S OCCUPATION BY RACE*

*For this tabulation, both full- and part-time workers are counted.
interview survey with workers.* In this tabulation, both full and part time workers were included; only full time workers were considered in the model estimation. Of the households tabulated, 19.3% (or 4078 households) were reported as being non-white, and 80.7% (or 17054) were reported as white.

Note that even in Washington, where non-discriminatory federal employment practices improve the job opportunities for non-whites, the white population held a disproportionate share of the higher income and higher status jobs in 1968. This fact, along with other socioeconomic differences between white and non-white households, has important implications for making inferences about differences in preferences from variation in utility function parameter estimates of the two groups. Errors in specification, particularly the omission of variables which vary across racial groups, may well show up in the value of utility function estimates of the coefficients of variables which are in some way correlated with the omitted variables. If behavioral conclusions drawn from the model results are to be valid, great care must be exercised in the model specification.

Table 6.2 is a tabulation of the distribution of household income for "blue collar" and "white collar" primary workers in the entire survey sample. In this tabulation, blue collar was defined as all occupations

---

*The rankings on the 0 to 9 scale in Table 6.1 were used to define the primary worker in households with more than one full time worker. In the case of a tie involving the head of the household, the head was selected; otherwise a random decision rule was applied.
<table>
<thead>
<tr>
<th>Income ($/yr.)</th>
<th>Blue Collar</th>
<th>White Collar</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than 3000</td>
<td>503</td>
<td>104</td>
<td>607</td>
</tr>
<tr>
<td>3000 - 3999</td>
<td>428</td>
<td>95</td>
<td>523</td>
</tr>
<tr>
<td>4000 - 5999</td>
<td>1165</td>
<td>463</td>
<td>1628</td>
</tr>
<tr>
<td>6000 - 7999</td>
<td>1613</td>
<td>1342</td>
<td>2955</td>
</tr>
<tr>
<td>8000 - 9999</td>
<td>1399</td>
<td>2009</td>
<td>3408</td>
</tr>
<tr>
<td>10,000 - 11,999</td>
<td>993</td>
<td>2324</td>
<td>3317</td>
</tr>
<tr>
<td>12,000 - 14,999</td>
<td>669</td>
<td>2567</td>
<td>3236</td>
</tr>
<tr>
<td>15,000 - 19,999</td>
<td>389</td>
<td>2334</td>
<td>2723</td>
</tr>
<tr>
<td>20,000 - 24,999</td>
<td>112</td>
<td>1146</td>
<td>1258</td>
</tr>
<tr>
<td>25,000 or more</td>
<td>55</td>
<td>849</td>
<td>904</td>
</tr>
<tr>
<td>No Answer</td>
<td>25</td>
<td>36</td>
<td>61</td>
</tr>
<tr>
<td>TOTAL (except no answer)</td>
<td>7326</td>
<td>13233</td>
<td>20559</td>
</tr>
<tr>
<td>GRAND TOTAL</td>
<td>7351</td>
<td>13269</td>
<td>20620</td>
</tr>
</tbody>
</table>

Table 6.2
DISTRIBUTION OF HOUSEHOLD INCOME BY OCCUPATION
(with column percentages of those responding)
with codes 4 or less in Table 6.1. For this reason, the "blue collar" and "white collar" terms are somewhat misleading, since the decision is largely based on the level of skill associated with the job rather than the type of occupation. Using this definition, the blue collar and white collar households constitute 36% and 64% of the households with workers respectively.

Income, tabulated in Table 6.2, obviously plays a significant role in determining mobility decisions. An often quoted "rule of thumb" is that a household expends approximately 25% of its income on housing (Grigsby, 1967). Income also determines the type of alternatives which a household will perceive as being available. For example, households at the low end of the income scale are very unlikely to perceive owning many automobiles or owning a single family home as feasible options. Conversely, high income households frequently have more than one car and own their homes.

Two other important socioeconomic variables, household size and the number of full-time workers, are tabulated in Table 6.3.* This table further indicates the relative importance of understanding the mobility decision of multiple and zero worker families. The classic economic model of a single worker household applies to less than 60% of the surveyed households; multiple and zero worker households constitute roughly 30% and 10% of the population respectively.

Furthermore, less than half of the multiple worker families have no non-working members. This provides strong evidence for the argument that

---

*This table, extracted from Cambridge Systematics (1974), excludes all households headed by someone 65 or older.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7-9</th>
<th>10+</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>23532</td>
<td>4085</td>
<td>5734</td>
<td>4374</td>
<td>4288</td>
<td>2833</td>
<td>1350</td>
<td>942</td>
</tr>
<tr>
<td></td>
<td>* 17.4</td>
<td>24.4</td>
<td>18.6</td>
<td>18.2</td>
<td>11.2</td>
<td>5.7</td>
<td>4.0</td>
<td>.5</td>
</tr>
<tr>
<td>NO ANSWER</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FULL TIME WORKERS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>2443</td>
<td>1212</td>
<td>710</td>
<td>209</td>
<td>140</td>
<td>66</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>* 49.6</td>
<td>29.1</td>
<td>8.6</td>
<td>5.7</td>
<td>2.7</td>
<td>2.0</td>
<td>2.0</td>
<td>.3</td>
</tr>
<tr>
<td>1</td>
<td>14064</td>
<td>2873</td>
<td>2615</td>
<td>2488</td>
<td>2817</td>
<td>1770</td>
<td>868</td>
<td>564</td>
</tr>
<tr>
<td></td>
<td>* 20.4</td>
<td>18.6</td>
<td>17.7</td>
<td>20.0</td>
<td>12.6</td>
<td>6.2</td>
<td>4.0</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>6035</td>
<td>1389</td>
<td>1037</td>
<td>616</td>
<td>329</td>
<td>227</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* 39.8</td>
<td>23.0</td>
<td>17.2</td>
<td>10.2</td>
<td>5.5</td>
<td>3.8</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>799</td>
<td>288</td>
<td>211</td>
<td>143</td>
<td>74</td>
<td>75</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>* 36.0</td>
<td>26.4</td>
<td>17.9</td>
<td>9.3</td>
<td>9.4</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>163</td>
<td>83</td>
<td>28</td>
<td>25</td>
<td>18</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>* 51.0</td>
<td>17.2</td>
<td>15.3</td>
<td>11.0</td>
<td>5.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>10</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>* 45.5</td>
<td>13.6</td>
<td>31.8</td>
<td>9.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 OR MORE</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>* 33.3</td>
<td>16.7</td>
<td>50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.3**

NUMBER OF WORKERS VS. HOUSEHOLD SIZE WITH ROW PERCENTAGES AND MEANS
second workers such as married women do not necessarily leave the labor force when they have children. Categorizing such workers as marginal to the household's location decision may prove misleading. Since 1968 it is likely that there are still more multiple worker households, and the need to better understand their mobility decisions has become if anything greater over time.

The next set of tables explore the distribution of mobility choices within the full cluster sample of households. The first of these, Table 6.4, presents the distribution of housing choice. As discussed in the preceding chapter, households residing in two-family structures were eliminated, and single family detached and single family attached structures were grouped into a single category.

As is typical in most American cities, the number of renters and owners is nearly evenly divided. The households reporting that they owned a garden-style or walk-up apartment are presumably condominium owners or owner-occupiers of multi-family dwellings. Because this group constitutes only .7% of the sample and since dwelling unit value information is not reported in the census for this type of unit, all households in this category were eliminated from the sample used for estimation. Furthermore, the alternative of owning a garden-style or walk-up apartment was not included in the set of available alternatives in the actual model estimation.
<table>
<thead>
<tr>
<th></th>
<th>Single Family</th>
<th>Garden-Style or Walk-up</th>
<th>High Rise</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own</td>
<td>2393 49.4%</td>
<td>37 .7%</td>
<td>0 0%</td>
<td>2430 50.2%</td>
</tr>
<tr>
<td>Rent</td>
<td>694 14.3%</td>
<td>1228 25.3%</td>
<td>493 10.2%</td>
<td>2415 49.8%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3087 63.7%</td>
<td>1265 26.1%</td>
<td>493 10.2%</td>
<td>4845 100.0%</td>
</tr>
</tbody>
</table>

Table 6.4

HOUSING TYPE CHOICE IN CLUSTER SAMPLE (with sample percentages*)

*percentages may not total due to roundoff
Tables 6.5 and 6.6 present the automobile ownership and mode to work for the primary worker of renters and owners respectively. Households reporting owning garden-style apartments are not included in this and subsequent tabulations. As one would anticipate, households choosing to own their homes have much higher automobile ownership and take the car mode to work much more frequently than do renters. The average auto ownership among owners is 1.56 autos per households, while the corresponding figure for renters is only 1.06.

This observed correlation between choice of housing tenure and auto ownership provides at least some limited support to the hypothesis that they are determined by some interdependent process. One might hypothesize that the mobility decisions of owning a dwelling unit and owning more than one auto are related in a way which makes them jointly more desirable than either taken independently. Similarly, renting and low auto ownership may be related in a synergistic way. At this level of analysis, however, these conclusions must be considered as at best tentative possibilities. Perhaps the observed correlations result from differences in the socio-economic characteristics of renters and owners rather than any actual interdependency in the choice process. The hypothesis of interdependence among specific dimensions of the mobility bundle is explored further in the following chapter on model specification and estimation.

Tables 6.7 and 6.8 present the same tabulations as in Tables 6.5 and 6.6 but for households in single and multiple family dwelling respectively.
<table>
<thead>
<tr>
<th></th>
<th>Auto Driver</th>
<th>Auto Passenger</th>
<th>Transit</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Autos</td>
<td>18</td>
<td>28</td>
<td>48</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>.7%</td>
<td>1.2%</td>
<td>2.0%</td>
<td></td>
</tr>
<tr>
<td>One Auto</td>
<td>416</td>
<td>321</td>
<td>122</td>
<td>859</td>
</tr>
<tr>
<td></td>
<td>17.4%</td>
<td>13.4%</td>
<td>5.1%</td>
<td></td>
</tr>
<tr>
<td>Two or more Autos</td>
<td>1063</td>
<td>336</td>
<td>41</td>
<td>1440</td>
</tr>
<tr>
<td></td>
<td>44.4%</td>
<td>14.0%</td>
<td>1.7%</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>1497</td>
<td>685</td>
<td>211</td>
<td>2393</td>
</tr>
<tr>
<td></td>
<td>62.6%</td>
<td>28.6%</td>
<td>8.8%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.5
AUTOMOBILE OWNERSHIP VS. MODE OF PRIMARY WORKER FOR OWNERS
(with sample percentages*)

*percentage totals may not sum correctly due to roundoff
<table>
<thead>
<tr>
<th></th>
<th>Auto Driver</th>
<th>Auto Passenger</th>
<th>Transit</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Autos</td>
<td>38</td>
<td>125</td>
<td>338</td>
<td>501</td>
</tr>
<tr>
<td></td>
<td>1.6%</td>
<td>5.2%</td>
<td>14.0%</td>
<td>20.7%</td>
</tr>
<tr>
<td>One Auto</td>
<td>785</td>
<td>346</td>
<td>208</td>
<td>1339</td>
</tr>
<tr>
<td></td>
<td>32.5%</td>
<td>14.3%</td>
<td>8.6%</td>
<td>55.4%</td>
</tr>
<tr>
<td>Two or more Autos</td>
<td>433</td>
<td>112</td>
<td>30</td>
<td>575</td>
</tr>
<tr>
<td></td>
<td>17.9%</td>
<td>4.6%</td>
<td>1.2%</td>
<td>23.8%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1256</td>
<td>583</td>
<td>575</td>
<td>2415</td>
</tr>
<tr>
<td></td>
<td>52.0%</td>
<td>24.1%</td>
<td>23.8%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 6.6
AUTOMOBILE OWNERSHIP VS. MODE OF PRIMARY WORKER FOR RENTERS (with sample percentages*)

*percentage totals may not sum correctly due to roundoff
<table>
<thead>
<tr>
<th></th>
<th>Auto Driver</th>
<th>Auto Passenger</th>
<th>Transit</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Zero Autos</strong></td>
<td>26</td>
<td>53</td>
<td>105</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>.8%</td>
<td>1.7%</td>
<td>3.4%</td>
<td>6.0%</td>
</tr>
<tr>
<td><strong>One Auto</strong></td>
<td>595</td>
<td>411</td>
<td>175</td>
<td>1181</td>
</tr>
<tr>
<td></td>
<td>19.3%</td>
<td>13.3%</td>
<td>5.7%</td>
<td>38.3%</td>
</tr>
<tr>
<td><strong>Two or more Autos</strong></td>
<td>1285</td>
<td>384</td>
<td>53</td>
<td>1722</td>
</tr>
<tr>
<td></td>
<td>41.6%</td>
<td>12.4%</td>
<td>1.7%</td>
<td>55.8%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>1906</td>
<td>848</td>
<td>333</td>
<td>3087</td>
</tr>
<tr>
<td></td>
<td>61.7%</td>
<td>27.5%</td>
<td>10.8%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 6.7

AUTOMOBILE OWNERSHIP VS. MODE OF PRIMARY WORKER FOR HOUSEHOLDS IN SINGLE FAMILY DWELLINGS (with sample percentages*)

*percentage totals may not sum correctly due to roundoff
<table>
<thead>
<tr>
<th></th>
<th>Auto Driver</th>
<th>Auto Passenger</th>
<th>Transit</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Autos</td>
<td>30</td>
<td>100</td>
<td>281</td>
<td>411</td>
</tr>
<tr>
<td></td>
<td>1.7%</td>
<td>5.8%</td>
<td>16.3%</td>
<td>23.9%</td>
</tr>
<tr>
<td>One Auto</td>
<td>606</td>
<td>256</td>
<td>155</td>
<td>1017</td>
</tr>
<tr>
<td></td>
<td>35.2%</td>
<td>14.9%</td>
<td>9.0%</td>
<td>59.1%</td>
</tr>
<tr>
<td>Two or more</td>
<td>211</td>
<td>64</td>
<td>18</td>
<td>293</td>
</tr>
<tr>
<td>Autos</td>
<td>12.3%</td>
<td>3.7%</td>
<td>1.0%</td>
<td>17.0%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>847</td>
<td>420</td>
<td>454</td>
<td>1721</td>
</tr>
<tr>
<td></td>
<td>49.2%</td>
<td>24.4%</td>
<td>26.4%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 6.8

AUTOMOBILE OWNERSHIP VS. MODE OF PRIMARY WORKER
FOR HOUSEHOLDS IN MULTI-FAMILY DWELLINGS
(with sample percentages*)

*percentage totals may not sum correctly due to roundoff
The results are quite analagous. Households choosing single family dwellings choose multiple auto ownership and take the car mode to work far more frequently than do households in multi-family dwellings. As before, this correlation does not imply causality. However, it does establish some basis for further exploring possible relationships between mobility choices.

6.6 Summary of Chapter 6

As the field of disaggregate modelling progresses, many new data sets specifically designed for disaggregate estimation will probably become available. Such data collections efforts are already underway on a pilot basis. However, at the present time, data for estimating the types of multi-dimensional mobility choice models such as those developed in this study must be pieced together from a variety of potentially inconsistent sources.

This chapter described the key assumptions made in the course of building an appropriate data base. Three primary sources, the Washington D.C. 1968 home interview survey, highway and transit interzonal skim tree matrices, and the Census Fourth Count Housing A file, were used in this data gathering process. In addition, a variety of minor sources were used to obtain tax, urban service, and cost of living data.

In addition to describing the assumptions made in creating the data base, a number of cross-tabulations of both the demographic characteris-
tics of the Washington, D.C. population and the mobility decisions that population made were presented. These tabulations provide some limited support for the hypothesis that there is a great degree of interdependence among the aspects of the household's mobility decision.

The following chapter describes the mobility choice models estimated in the study, including the variables and specifications used and the parameter estimates. These models attempt to capture the joint behavioral process of mobility choice to the greatest extent feasible within the limitations of the available data.
7.1 Specification of Behavioral Models

In the preceding chapters, a comprehensive theory of mobility and travel choice was proposed, a variety of causal mechanisms underlying mobility choice were considered, and an appropriate data set for the model estimation was described. However, the transition from the elegance of analytic theory to the realities of actually specifying a behavioral model of a complex process such as the household mobility decision defines the boundary between science and art in econometric modelling. Even within the limits of available data, the number of possible model specifications is for practical purposes limitless. While analytic theory provides useful insights into the types of variables which might appear in the utility functions, it offers little information about how those variables should be measured or how they should interact with one another.*

This shortcoming of existing theory does not necessarily imply that theoretical model development is in itself useless. Rather, it implies that formal analytic theory should be coupled with a great deal of qualitative theory derived from prior studies and an intuitive understanding of the problem being considered.

*In a study of automobile ownership reviewed in Chapter 3, Burns, Golob, and Nicolaïdis derive a model specification directly from a formal theory. However, analytic tractability is achieved only with a great loss in the credibility of the behavioral content of the model formulation. In particular, the interactions among household members for use of available autos is greatly simplified.
Other factors aside from formal theory and informed judgement also enter into the choice of model specification. For example, the available estimation procedures for the multinomial logit model are restricted to utility functions which are linear in their parameters. Thus, certain non-linearities which might exist may be difficult or impossible to represent. Furthermore, many causal factors can be represented by such an enormous array of measures that it is computationally infeasible to include them all in the utility function. Decisions which can not be tested due to budget limitations must be made about which variables should be included and which can be safely omitted. Finally, the data itself imposes certain limitations on the specification. Some variables such as neighborhood or school quality are inherently difficult to measure. Frequently, a proxy measure such as the average income of residents or the local expenditure per pupil must be adopted, even though it reflects an input to a process which determines the causal variable rather than the output itself.

This chapter describes the judgements made in the specification of the mobility models for single and multi-worker households and presents the model parameter estimates which were obtained in the course of the study. Separate models were estimated for single and multi-worker households, since their perception of alternative locations would seem likely to be governed by different processes. As might be anticipated, the results for single worker households are
far superior to those for households with more than one worker. This reflects the lack of a strong base of available theory about how multi-worker households make their mobility decisions combined with the severely limited computational resources available to this study, which made exploration of a full range of possible behavioral hypotheses impossible.

The description of the variables in the model is divided into two sections. First, the measures used to represent each of the categories of variables discussed in Chapter 2 are presented, without reference to the way in which those measures are structured as independent variables in the utility functions. Then, in a separate section, the actual structure of the utility functions is described. For the sake of clarity, these sections focus on the specification of the single worker household model; all changes made for multi-worker households are discussed in the description of the estimation results for those models.

Following this discussion, the methods used to determine which of the enormous set of possible alternatives were available to each household is considered. A set of rules reflecting data availability and behavioral factors is proposed and discussed.

Estimation results for two types of models are then described. In each household group, single and multiple worker, the first model presented is a conditional location and housing choice model, where
automobile ownership and mode to work are taken as given. These models were estimated in order to explore various specifications of the location and housing aspects of the household utility function without incurring the substantial computational cost associated with estimating the complete joint model.* Ben-Akiva (1973) has shown that in theory, the estimation of conditional multinomial logit models provides consistent estimates of the parameters of variables in the utility function which change their value across the subset of alternatives or choice groups considered.** Obviously, the coefficients of variables which are constant across all alternatives in the conditional model are unidentified, since for each individual household those variables do nothing to distinguish one alternative from another.

Following the description of the estimation results for the conditional location and housing models, the estimates for the corresponding joint model are described.

Note that the discussion of the estimation results for each model presented is limited to a consideration of the signs and statistical properties of the estimated coefficients. The following chapter

---

* The automobile ownership and mode to work aspects of the mobility decision are extensively considered in the previously discussed auto ownership models by Lerman and Ben-Akiva (1975).

** McFadden and Reid (1975) describe conditions under which some conditional logit models might provide consistent parameter estimates even though estimation of the joint model does not. In particular, this may occur when the disturbance term of the conditional utility function has the assumed Weibull distribution while the disturbance for the joint does not.
includes a comparison between the models and considers in detail the behavioral implications of the results.

7.2 Measures in the Joint Model Specification

In Chapter 2, six classes of variables which affect the mobility choice were defined as follows:

1) transportation level of service to work;
2) automobile ownership attributes;
3) locational attributes;
4) housing attributes;
5) spatial opportunities;
6) socioeconomic characteristics.

The measures used to model the effect of each of the classes will be considered below.

The transportation level of service to work can be measured both for car and for transit in a number of ways. Traditional mode choice models have generally focused on the most directly measurable trip attributes: in-vehicle time, out-of-vehicle time, and travel cost. More recently, researchers have attempted to introduce attitudinal measures which can reflect comfort, convenience, etc. (Tardiff, 1975)

Such attitudinal measures may prove to be significant determinants of the mobility decision, since, for example, shifts in the amenities associated transit travel may greatly enhance the attractiveness of locations near transit stations. However, aside from the lack of attitudinal information in the Washington data, current under-
standing of the appropriate use of attitudinal information in choice modelling would appear to be too poor to make a departure from traditional measures warranted. Furthermore, as with the use of perceived levels of service data discussed in Chapter 6, the forecasting of attitudinal data presents major obstacles which, while potentially surmountable, will require extensive research in the future.

In addition to the three traditional transportation level of service measures, a dummy variable which reflects the additional disutility associated with using the car mode to travel to a downtown workplace was introduced. This variable, the significance of which has been empirically verified in previous studies by Cambridge Systematics (1974), measures the effect of the frustration associated with downtown congestion in the central business district and the high variance of car travel time associated with downtown oriented trips.

**Automobile ownership attributes** can be measured by a number of variables. However, most of these measures such as vehicle size, horsepower, age, and make are specific to the type of car under consideration, while the mobility choice models to be estimated deal only with the number of autos as the available alternatives. Automobile ownership attributes must therefore be greatly simplified. For this reason, only the average annual cost of auto ownership (not including location dependent factors such as insurance or tags) was used in the model. This value was assumed to be $800 per auto.* The benefits

---

*This figure is a modification of one used by Lerman and Ben-Akiva (1975) and Burns, Golob and Nicolaidis (1975), corrected for insurance costs, which are separated here into a distinct measure.
which a household derives from owning automobiles are reflected in
the structure of the spatial opportunity variables and the set of
mobility alternatives open to a household when autos are owned.

A broad range of location attributes was discussed in the theore-
etical development presented in Chapter 4. The factors proposed
there included urban services, taxes, pollution and the racial com-
position and quality of the neighborhood surrounding a particular
residential site. Previous studies such as those described in the
literature review in Chapter 3 have measured these attributes in a
number of ways.

For this study, the effect of pollution sources was entirely
ignored, both due to the lack of available data and the assumption
that the most significant forms of pollution which affect mobility
decisions are highly localized noise and visual pollution. The
effect of these types of pollution, when averaged at the tract level,
would in any case be very difficult to measure in small samples.

Local urban services can be measured in terms of the entire
spectrum of municipal services such as schools, police, fire-
fighting, recreational facilities, sanitation, sewerage, road main-
tenance, etc. However, one service, the public schools, has been
gen generally viewed as being of particular significance in the evalua-
tion of location decisions. Education accounts for roughly half of
all local government expenditures in the United States and is an
important indicator of both the quality of other services and the
attitudes of the community.
For this reason, annual per pupil school expenditure was used in the model. However, in the District of Columbia, many residents use the extensive private school system, particularly upper status residents who probably view school quality as extremely important and can afford the rather expensive alternative to public schools. This factor was accounted for by not including school expenditures in the utility function for sites in the District of Columbia and by introducing an extra dummy variable for District of Columbia locations to correct for this.

Urban services are only one aspect of the effect of local government on locational decisions; the other is local taxes. Taxes imposed on real property for home owners by municipalities (as a function of property values) were included in the model for the owner-occupied single family dwelling alternatives. In addition, state income taxes as a function of household size, marital status and income were estimated from standard tax formulas.

Other costs which vary across locations (as well as levels of auto ownership) such as automobile insurance tags and personal property taxes were also used in the model. However, incidental expenses with only a minor location dependent component such as utility costs and sales taxes were ignored.

Racial composition is of particular significance in the Washington area, where the District of Columbia is predominantly black and the
surrounding suburbs are predominantly white. For this reason, the 
fraction of non-white households in a location was used in the model 
specification. Obviously, this measure is perceived differently by 
white and non-white households, and this difference must be reflected 
in the structure of the utility functions.

Neighborhood quality is a particularly difficult attribute to 
measure, since it is a generic term for a complex bundle of attributes 
and may be perceived in many ways by different households. The measure 
ultimately selected is the squared difference between the household's 
income and the average tract income. One would expect that all else 
being equal, people would prefer not to live in an area populated 
primarily with lower income residents. To some extent, they might prefer 
to live among those with higher incomes. However, at some point one 
might speculate that a household would rather not live in an area where 
its income was insufficient to maintain the life style of its neighbors. 
Experimentation with a number of forms of the income differential measure 
lead to the conclusion that the desire not to live with those of lower 
income was quite strong, while the reverse effect was ambiguous and at 
best quite weak and difficult to measure reliably in small samples. This 
result is reflected in the utility specification described in Section 7.3.

The final locational attribute used in the specification is the net 
residential density of the location. This measure in part represents the
general character of the neighborhood in which a particular housing unit is located. In addition, it is a measure of average lot size, which is a housing rather than a locational attribute.

The fourth class of variables, housing attributes, is perhaps the most poorly represented group. Aside from a number of dummy variables for housing types and the density measure discussed above, the only housing attribute used was the annual housing cost in each location. For rental housing, this was measured as the annual gross rent, estimated by the U.S. Census.* For owner-occupied single family dwellings, it was assumed that the total annual cost of any unit was 12% of the house's value (Ingram, et al., 1972). This cost is exclusive of property taxes described previously as a locational attribute.

The fifth type of variable, spatial opportunities, are perhaps the most difficult class of variables to represent and measure.** Since they are a composite of all non-work trips, some way of combining the characteristics of various possible trips with the relative likelihood of the household making those trips must be found.

Typical measures of accessibility which have been used in empirical studies are generally of two types. The first approach is a weighted

---

*Gross rent differs from the usual contract rent in that it includes the cost of utilities and other expenses where they are not part of the monthly payment to the landlord.

**This discussion of spatial opportunities is adapted from Cambridge Systematics (1974).
averaging method, which typically takes the following form:

\[
\text{Acc}_{im}^k = \sum_j A_j^k f^k(t_{ijm})
\]

where: \(\text{Acc}_{im}^k\) is the accessibility of location \(i\) to opportunity \(k\) by mode \(m\);

\(A_j^k\) is some measure of the attractiveness of place \(j\) to opportunity \(k\);

\(t_{ijm}\) is a level of service measure from \(i\) to \(j\) by mode \(m\);

\(f^k(t_{ijm})\) is a discrete friction factor table or a continuous function such as: \(f^k(t_{ijm}) = t_{ijm}^\alpha\).

The level of service measures, \(t_{ijm}\), used in this type of variable can either be single trip characteristics such as wait time and cost, or combinations of measures which collapse a vector of attributes into a single variable. The latter is generally referred to as a generalized price, and has been applied in a number of studies (Ben-Akiva, 1973; CRA, 1972).

The second type of spatial opportunity variable has been used by Dunphy and Wickstrom (1972) and Sherman, et. al. (1974). This approach is based on an isochron analysis describing the number of opportunities within a certain level of service, denoted as \(\hat{t}\). Expressed mathematically,

\[
\text{Acc}_{im}^k = \sum_j A_j^k \delta_j
\]

*For examples of this type of variable, see TRC(1967), Shindler and Ferreri (1967), Brigham (1965) and Bradbury, et. al. (1974)
where:
\[ \delta_j = \begin{cases} 
1 & \text{for } t_{ijm} \leq \hat{t}_{ijm} \\
0 & \text{for } t_{ijm} > \hat{t}_{ijm}
\end{cases} \]

Both these measures have serious shortcomings in a truly behavioral model. The first measure requires a specific arbitrarily selected attraction measure for weighting the alternative destinations to which a household might travel. The second is only a characteristic of the supply of transportation and bears no direct causal relation to the non-work trips a household will make.

The approach used in this study relies on a much more behavioral means of combining the level of service to different non-work destinations. It is based on the following line of reasoning.

When a household makes its mobility decision, it has not yet determined its pattern of non-work travel. However, given any choice of a mobility bundle, one would then be able to determine the probability with which it would travel to each destination. These probabilities themselves depend on the transportation level of service and the attractiveness of each destination, but they also depend on the characteristics of the household. Therefore, it makes theoretical sense to use an estimate of the household level probabilities to weight the level of service for non-work travel.

Expressed mathematically, this spatial opportunity measure is determined as follows:*

*See CRA (1972) and Ben-Akiva (1973) for a more detailed discussion of this type of composite variable.
\[ Acc_{im}^k = \sum_j t_{ijm}^P (i,j,m), \]

where:

\[ P (i,j,m) \] is the probability of the household travelling from \( i \) to \( j \) by mode \( m \).

Actually measuring this type of accessibility gives rise to a number of practical problems. First, what types of spatial opportunities should be used? It was decided that the most relevant non-work travel purpose is shopping, and that other spatial opportunities play a secondary role and can be ignored.* This greatly reduced the computational problem of creating the accessibility measures without substantially sacrificing important causal effects.

The second question is what measure of level of service is most appropriate. While it is possible to use each measure separately, this gives rise to an unwieldy number of variables. Hence, generalized prices which are a weighted linear function of in-vehicle time, out-of-vehicle time and out-of-pocket cost were used. Furthermore, the value of time used was a function of income, reflecting the hypothesis that high income households would be willing to pay more to save time than low income households. Thus,

\[ t_{ijm} = \alpha_1 OVTT_{ijm} + \alpha_2 IVTT_{ijm} + \alpha_3 (OPTC_{ijm}/Y) \]

*As hypothesized in section 4.8, multi-worker households may perceive employment spatial opportunities as relevant in their location decision. This is explored in the description of the multi-worker household models in Section 7.7 and 7.8.
where:

\[ Y = \text{household annual income;*} \]

\[ OVTT_{ijm} = \text{out-of-vehicle travel time from i to j by mode m;} \]

\[ IVTT_{ijm} = \text{in-vehicle travel time from i to j by mode m;} \]

\[ OPTC_{ijm} = \text{out-of-pocket travel cost from i to j by mode m;} \]

\[ \alpha_1, \alpha_2, \alpha_3 \] \text{are constants.}

Note that all these level of service measures are for shopping trips, which are typically made in the off-peak hours. Thus, they are substantially different from the work trip level of service measures.

The last two questions that arise from the use of this type of accessibility measure is how can the probabilities, \( P(i,j,m) \) and the parameters of the generalized prices, \( \alpha_1, \alpha_2, \) and \( \alpha_3, \) be estimated. Since the underlying motivation for using this type of variable is its behavioral, disaggregate interpretation, it seems logical that a previously calibrated disaggregate choice model should be used.

This is precisely what was done in this study. Using a model developed by Ben-Akiva (1973), the probabilities of every household in the sample selecting each shopping alternative were determined based on off-peak level of service data.** Ben-Akiva's model is in itself a joint choice model; in this case, choice of mode and destination are

---

*Actually, the coefficient \( \alpha_3 \) was estimated with \( Y \) measured as a 1 through 10 code, so in computing \( \tau_{ijm} \) the same code was used.

**Zone to district skim trees were developed from zone to zone trees in order to maintain consistency with Ben-Akiva's original estimation methodology. These were then converted to the tract level using representative zones.
considered. Using the joint probability of destination and mode, $P(d,m)$, the conditional probabilities, $P(d|\text{car})$ and $P(d|\text{transit})$ were derived.

The parameters $\alpha_1$, $\alpha_2$, and $\alpha_3$ were taken from the utility function of this joint probability model. The actual values used are as follows:

$\alpha_1 = -.0633$

$\alpha_2 = -.0164$

$\alpha_3 = -.0757$.

In order to adjust the units of the variable to be generalized costs instead of utility, the value was multiplied by minus one.

Using the forecasted probabilities, the attributes for the "expected shopping trip" by both auto and transit were determined. These values are a function of both the location under consideration and the income of the household, and hence are highly disaggregate measures of spatial opportunities.

The final group of variables, socioeconomic characteristics, are a special type of measure in that they do not vary across alternatives in a household choice set. Thus, in the specification of the utility function they must be transformed by either combining them with other variables or allowing them to have different coefficients in different utilities (i.e. allowing them to be "alternative specific").

Because the mobility choice process is so complex, the number of socioeconomic characteristics which must be considered is quite large. Simple consumer theory indicates that income after taxes should appear
in the utility function.* In addition, a household's stage in its life
cycle alters its perception of many of the previously discussed vari-
able. For example, households without children may not care about
local school quality, since they do not benefit from it. Large house-
holds may have greater need for the space associated with single family
dwellings, while smaller ones might desire apartments. Households with
large numbers of licensed drivers may desire to own more automobiles,
though the chances of any worker using one of the cars available to
travel to work probably decreases when there are more drivers. In
multiple worker households, particularly where some workers are second-
ary to the household location decision, the importance of accessibility
to employment may depend on how many workers there are. As discussed
previously, a household's response to the racial composition of a neigh-
borhood depends on its own race. Finally, federal and state
income taxes depend on the marital status of the head of the
household.

Each of the measures discussed above are used in the specification
of the joint mobility choice utility functions defined in the following
section. Table 7.1 presents a summary of the measures used by category.

*Recall that in the context of choice theory the term utility function
is actually an indirect utility function, and therefore includes income
and prices rather than simply the quantity and attributes of various
goods consumed.
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation Level of Service to Work</td>
<td>in-vehicle time, out-of-vehicle time (for car and transit), car operating cost, transit fare, a CBD workplace dummy variable</td>
</tr>
<tr>
<td>Automobile Ownership Attributes</td>
<td>assumed cost of $800/auto</td>
</tr>
<tr>
<td>Locational Attributes</td>
<td>per pupil school expenditures, state income taxes, real and personal property taxes, auto insurance and tag costs, fraction non-white, difference between average tract income and household income, net residential density</td>
</tr>
<tr>
<td>Housing Attributes</td>
<td>gross rents, housing values</td>
</tr>
<tr>
<td>Spatial Opportunities</td>
<td>generalized shopping price by car and transit</td>
</tr>
<tr>
<td>Socioeconomic Characteristics</td>
<td>income (after federal income taxes), marital status, household size, number of licensed drivers, existence of school-age children in household, race</td>
</tr>
</tbody>
</table>

Table 7.1
SUMMARY OF MEASURES USED IN THE MODEL
7.3 Specification of the Joint Utility Function

Given the measures discussed in the previous section there are a virtually limitless number of possible utility function specifications which might be explored. This section describes the chosen joint model used for single worker households; changes made for the specification of multi-worker households are described in section 7.7.

In the joint model there are an extremely large number of possible location/housing/auto ownership/mode to work combinations. Thus, the number of utility functions is correspondingly great. Rather than consider each utility function individually, every variable will be defined as pertaining to all alternatives but as taking zero value for those utilities where it is not included.

The first group of variables are constant terms in the utility function. These constants measure the so-called "pure alternative" effects, i.e. the net effect of all attributes of an alternative which are not measured by the other variables. In theory, a constant could be introduced into all but one utility function, which acts as a base against which the effect of the other variables is measured. The choice of the base utility function is arbitrary and has no effect on the parameter estimates of other variables or the choice probabilities. In practice, however, alternatives such as locations which are unranked and very numerous do not have constant associated with them unless they have particular attributes which make them distinguishable, as with the central business district in models of destination choice.
Even ignoring the location choice group, the number of possible options is quite large. A household has a maximum of two modes: car and transit; three auto ownership levels: zero, one, and two or more; and four housing types: own a single family house, rent a single family house, rent a garden style or walk-up apartment and rent a high rise apartment. After eliminating the logically inconsistent alternative of zero auto ownership and car to work, there are twenty possible options.*

In order to limit the number of dummy variables to something less than nineteen, (the number of possible options minus one for a base), some way of approximating each independent effect by a linear combination of a smaller number must be found. While Nerlove and Press (1973) explore a technique which uses pairwise associations among alternatives to test for relationships among the options, a much simpler approach was adopted here.** Each choice group was given a constant term for all its members but one, and some of the interactions among choice groups that the exploratory data analysis in Chapter 6 indicated as being significant were assigned constants. The resulting set of eight constant terms are as follows:

*Not all these options are necessarily available to every household or are available in every location, as will be discussed in the next section.

**Nerlove and Press construct an odds ratio equal to $N_{ij}N_{jj}/N_{ij}N_{ji}$, where $N_{ij}$ is the number in the sample choosing the $i$th option from one choice group and the $j$th from another. Values of this statistic near unity imply independence.
DRENT1 = \begin{cases} 
1 & \text{in the rent single family dwelling alternatives} \\
0 & \text{otherwise} 
\end{cases}

DRENTG = \begin{cases} 
1 & \text{in the rent garden style or walk-up apartment} \\
0 & \text{otherwise} 
\end{cases}

DRENTH = \begin{cases} 
1 & \text{in the rent high rise apartment alternatives} \\
0 & \text{otherwise} 
\end{cases}

DA01 = \begin{cases} 
1 & \text{in the one auto alternatives} \\
0 & \text{otherwise} 
\end{cases}

DA02 = \begin{cases} 
1 & \text{in the two or more auto alternatives} \\
0 & \text{otherwise} 
\end{cases}

DCAR = \begin{cases} 
1 & \text{in the car to work alternative} \\
0 & \text{otherwise} 
\end{cases}

DAPTSTYL = \begin{cases} 
1 & \text{in the rent garden style, walk-up or high rise apartment and own less than two autos alternatives} \\
0 & \text{otherwise} 
\end{cases}

DSUBSTYL = \begin{cases} 
1 & \text{in the own single family dwelling and own two or more autos alternatives} \\
0 & \text{otherwise} 
\end{cases}

The next two variables represent the travel time aspects level of service to work. Traditionally in mode choice studies, these variables have been expressed as simply the in-vehicle and out-of-vehicle time. More recent work by Koppelman (1975) and Cambridge Systematics (1974) has indicated that the disutility of out-of-vehicle time may be perceived as a function of the total trip length, which can be measured by travel distance. After experimentation with a number alternative functional forms, the specification used by Koppelman was chosen. This is as follows:
TOTIME = total two way travel time (in minutes)

OVTT/DIST = \frac{two way out-of-vehicle time (in minutes)}{two way travel distance (in miles)}

In addition to these variables, a dummy variable was defined to reflect the added disutility associated with the use of a car in the downtown as follows:

\[
DCITY = \begin{cases} 
1 & \text{for households with downtown workplaces in the car to work alternatives} \\
0 & \text{otherwise} 
\end{cases}
\]

The next variable arises from the fact that there are a large number of monetary measures in the model, including household income, federal, state and local taxes, housing costs, auto ownership costs, and out-of-pocket travel costs for the work trip. Clearly, one would like to avoid introducing a separate variable for each of these cost factors. The question is how can these attributes be combined into a single variable representing the money which would be available to the household if it selected each alternative. This was done by formulating a variable, termed for reference the Z variable, as follows:

\[
Z = Y - FTAX(Y,HHSIZE, MARRY) - STAX(Y,HHSIZE,MARRY,STATE) - PTAX*HVALUE - HOUSING - 800*AO - INS(STATE)*AO - 250*OPTC
\]

where:

\[
Y = \text{annual household income;}
\]

\[
FTAX(\ ) = \text{federal tax function;}
\]

\[
HHSIZE = \text{household size;}
\]

\[
MARRY = \text{marital status;}
\]
STAX ( ) = state tax function;
STATE = state of location;
PTAX = effective property tax rate;

\[ HVALUE = \begin{cases} 
\text{housing value in the own single family dwelling} & \text{alternative} \\
0 & \text{otherwise} 
\end{cases} \]

\[ HOUSING = \begin{cases} 
\text{annual gross rent for rented housing alternatives} & \text{alternative} \\
.12 \times HVALUE & \text{for own single family dwelling alternative} 
\end{cases} \]

AO = level of auto ownership in alternative;
INS ( ) = auto insurance, tags, and personal property taxes;
OPTC = daily two way work trip cost.

In words, the value of Z (in dollars per year) is an estimate of the amount of money a household has left after the following expenses:

1. federal taxes;
2. state taxes;
3. property taxes (if applicable);
4. housing cost;
5. direct auto ownership costs;
6. auto insurance, tags, and taxes;
7. commuting cost to work (250 annual work trips assumed).

The coefficient of the Z variable in the utility function should always be positive, reflecting the fact that all else being equal, households would rather have more money than less left for other things.

The intuitive arguments presented above can be formalized in the context of the classical consumer allocation problem in economics. Consider a household with income Y which must be allocated among N commodities \( x_1, x_2, \ldots, x_N \). The household has a utility function:

\[ U = U(x_1, x_2, \ldots, x_N). \]
Now suppose one is only interested in a subset of the N commodities, which for convenience will be taken to be the first K goods. Thus,

\[ U = U(x_1, x_2, \ldots, x_N) = U(x_1, x_2, \ldots, x_K, x_{K+1}, \ldots, x_N) \]

The budget constraint for the household can be written as:

\[ Y = p_1 x_1 + p_2 x_2 + \ldots + p_i x_i + \ldots + p_N x_N = \sum_{i=1}^{K} p_i x_i + \sum_{j=K+1}^{N} p_j x_j \]

where \( p_i \) and \( p_j \) are the prices of the \( i^{th} \) and \( j^{th} \) commodities respectively.

Now suppose we are willing to assume that the utility function is such that it can be represented as a function of two additively separable components, *i.e.*

\[ U = U(f_1(x_1, \ldots, x_K) + f_2(x_{K+1}, \ldots, x_N)) \]

where the commodities of interest appear in one function, \( f_1(\ ) \), and the remaining ones appear in the other, \( f_2(\ ) \).

Using this partition, the right hand term of the budget constraint, \( \sum_{j=K+1}^{N} p_j x_j \), is the amount of income expended on the second set of commodities. Now define the \( Z \) variable as that amount of income such that:

\[ Z = Y - \sum_{i=1}^{K} p_i x_i = \sum_{j=K+1}^{N} p_j x_j. \]

If all the prices of commodities \( x_{K+1}, \ldots, x_N \) remain constant, the utility function can be written as:

*This assumption of functionally separable utility functions is implicit in virtually all empirical analyses of consumer behavior. For a more detailed discussion, see Strotz (1957) and Goldman and Uzawa (1964).*
\[ U = U(f_1(x_1, \ldots, x_K) + f_2(Z)) \]

Thus, the Z variable stands as a proxy for the utility a household obtains from expenditures on the entire range of commodities which are not explicitly modelled.**

The use of the Z variable is a classic case of the selection of variables based on a priori reasoning; its justification is based on theoretical considerations rather than empirical ones. However, the motivation for developing this variable in the first place was that estimates of separate coefficients of the various cost components would be less reliable. By collapsing all costs into a single variable an extra piece of information has been added to the model formulation; it has been assumed that the marginal utility of any cost component is the same, regardless of what type of expenditure is considered. Stated more simply, it has been assumed that the household views an extra dollar as having the same value, regardless of where it is spent.

*The notation \( f_2 \) is used to indicate that implicitly the \( x_{K+1}, \ldots, x_N \) selected are such as to maximize the value of \( f_2(x_{K+1}, \ldots, x_N) \), subject to

\[ Z = \sum_{K+1}^{N} p_j x_j. \]

**A more formal and rigorous demonstration of this by deriving first order conditions for a utility maximum is possible.
It was decided that the Z variable should not enter the utility functions linearly; the utility a poor family derives from an extra dollar is probably much greater than that of a wealthy family. Thus, the marginal utility of money should decrease as the value of Z increases.* This hypothesis was reflected by using the natural log of Z as an independent variable rather than simply the value of Z.

The next variable used commonly appears in simple mode choice models in which auto ownership is assumed fixed. This variable is defined as:

\[
AALD = \begin{cases} 
  \text{number of autos in alternative} \\
  \text{number of licensed drivers in the household in the car to work alternatives} \\
  0 \text{ otherwise}
\end{cases}
\]

and represents the level of automobile availability which would be obtained if the household chose a given alternative. Alternatives with high auto availability should be associated with high car to work utilities relative to those for transit to work alternatives; hence, the expected sign of its coefficient is positive.

The next variable was designed to reflect another effect of the number of licensed drivers within a household. While the number of licensed drivers impacts on choice of mode to work through the AALD variable, it also should affect the level of auto ownership directly. The more licensed drivers in a household, the more likely it should be to select a high auto ownership level, independent of the mode to work.

*This hypothesis of decreasing marginal utility is commonly invoked in conventional consumer theory (Samuelson, 1970).
selected. This effect was measured by introducing a variable which reflects the number of licensed drivers into each utility function with a different coefficient for each auto ownership level except one, selected arbitrarily as a base. These variables were defined as follows:

\[
ILD_1 = \begin{cases} 
\frac{1}{\# \text{ of licensed drivers in the household}} & \text{for one auto alternative} \\
0 & \text{otherwise}
\end{cases}
\]

\[
ILD_2 = \begin{cases} 
\frac{1}{\# \text{ of licensed drivers in the household}} & \text{for two auto alternatives} \\
0 & \text{otherwise}
\end{cases}
\]

When these variables were originally introduced into the model, it was hypothesized that the effect for the two car alternative (as measured by the coefficient value) would be twice as great as the effect for the one car alternative. Statistical tests by Lerman and Ben-Akiva indicated that this was indeed the case, and for the models ultimately estimated \(ILD_1\) and \(ILD_2\) were combined into a single variable, \(ILD\), defined as follows:

\[
ILD = \begin{cases} 
0 & \text{for the zero auto alternatives} \\
ILD_1 & \text{for the one auto alternatives} \\
2ILD_1 & \text{for the two auto alternatives.}
\end{cases}
\]

The use of the inverse of the number of drivers rather than simply the number itself reflects the hypothesis that as the number of drivers increases, the marginal effect of an additional driver on the need for automobiles decreases. Clearly, the coefficient of \(ILD\) should be less than zero.

*The use of this variable is equivalent to estimating a model with \(ILD_1\) and \(ILD_2\) with the linear constraint that the coefficient of \(ILD_2\) is twice as large as the coefficient of \(ILD_1\).
Spatial opportunities influence the mobility decision in at least two ways. First, the absolute level of accessibility to shopping by either car or transit is probably important in a household's choice of location. Second, the level of shopping accessibility by car relative to that of transit affects the way in which the household will travel to shop, which in turn influences their desired level of automobile ownership.

The first of these effects was represented by a variable defined as

\[
GPTINV = \frac{1}{\text{expected generalized shopping price by transit}}
\]

This variable is zero when transit is completely unavailable since transit generalized price in such areas is for practical purposes infinite. One would expect that the coefficient of this variable would be positive, since decreased travel costs resulting from improved transit service should increase the household's utility.

Attempts to use a corresponding variable for the absolute level of car accessibility produced statistically insignificant coefficient estimates with an unexpected sign. This problem also occurred in a study by Wheaton described in Chapter 3. He attributes the result to the high levels of externalities (noise, traffic congestion, etc.) often associated with locations with good highway accessibility. Thus, the value of the car accessibility coefficient may be partially measuring the effect of some omitted variables. For this reason, it was not included in the final specification.
The effect of the relative accessibility was measured by a variable defined as follows:

$$R = \frac{\text{expected generalized car cost for shopping}}{\text{expected generalized transit cost for shopping}}$$

However, this variable does not change value for different auto ownership alternatives, and therefore must be introduced into the utility function as alternative specific.* Thus, the following two variables appear in the model:

$$R_1 = \begin{cases} R \text{ in one auto alternatives} \\ 0 \text{ otherwise} \end{cases}$$

$$R_2 = \begin{cases} R \text{ in two auto alternatives} \\ 0 \text{ otherwise} \end{cases}$$

As generalized shopping cost by car increases, the value of $R$ increases. One would anticipate that this increase in car cost would result in greater use of transit for shopping trips. Consequently, the likelihood of high auto ownership should decrease. To reflect this hypothesis, the coefficients of both $R_1$ and $R_2$ should be negative, since they both measure the effect of shopping accessibility relative to the zero auto-transit to work alternatives. Furthermore, the effect should be greater for the two auto alternatives than for the one auto options. This should result in the magnitude of the coefficient of $R_2$ being greater than that of $R_1$.

---

*This would not have been true had the shopping trip model used auto ownership as an explicit independent variable. However, it still might have been desirable to capture differences in the way accessibility is perceived for the one and two auto options.
In order to reflect the effect of household size on the desire for a living in single family dwelling, the following variable was defined.

\[
\text{HHSIZE1} = \begin{cases} 
\text{household size in single family dwelling alternative} \\
(\text{own or rent}) \\
0 \text{ otherwise}
\end{cases}
\]

The coefficient estimate of this variable should be greater than zero, since, all else being equal, larger households probably prefer single family dwellings more than do smaller households.

The next group of variables are all locational attributes as described in the preceding section. They are defined as follows:

\[
\text{INCDIFF} = \begin{cases} 
(Y - \overline{Y})^2 \text{ for } Y > \overline{Y}, \text{ where } \overline{Y} \text{ is the average annual tract income, (both } Y \text{ and } \overline{Y} \text{ are in thousands)} \\
0 \text{ otherwise}
\end{cases}
\]

\[
\text{FBFORW} = \begin{cases} 
\text{fraction of non-white households in tract for white households} \\
0 \text{ for non-whites}
\end{cases}
\]

\[
\text{FBFORB} = \begin{cases} 
\text{fraction of non-white households in tract for non-whites} \\
0 \text{ for non-whites}
\end{cases}
\]

\[
\text{DENSITY} = \text{net residential density in households per acre}
\]

\[
\text{SCHOOL} = \begin{cases} 
\text{per pupil school expenditures (in dollars per year)} \\
\text{except in District of Columbia} \\
0 \text{ in District of Columbia}
\end{cases}
\]

\[
\text{DOC} = \begin{cases} 
1 \text{ in District of Columbia} \\
0 \text{ otherwise}
\end{cases}
\]
The first variable is as discussed in section 7.2. The income differential is squared, reflecting the hypothesis that large differences are proportionately much more important than small ones. One would expect a negative coefficient for this variable. The opposite variable defined as non-zero when the households income is less than the average consistently yielded very small, statistically insignificant estimates with the wrong sign, and was omitted in the final specifications.

The two racial composition variables reflect the hypothesis that whites and non-whites perceive the racial composition quite differently. The coefficients of FBFORW and FBFORB should be negative and positive respectively.

The density variable is self-explanatory; a negative coefficient would be expected. The DOC dummy variable was defined to correct for the setting of the annual per pupil school expenditure variable to zero in the SCHOOL variable, the coefficient of which should have a positive sign. Note that SCHOOL is defined to be zero for households without children even though DOC is not defined this way. This was done to explore the possibility that the District has certain attributes which make it distinct from other locations regardless of whether or not a household has children.

The final variable in the model is the measure of tract size required to correct for the grouping of alternatives described in Chapter 5. This variable, denoted as ln $N_1$, is the natural logarithm of the
number of housing units of the type associated with the alternative in each tract. It should have a coefficient constrained to unity; however, both constrained and unconstrained models are reported.

Table 7.2 is a summary of the variable definitions. In order to derive the structure of the utility function for any particular location/housing/auto ownership/mode to work combination from the table, the variables which are set to zero for that alternative can be omitted. For example, if the twenty five variables in Tables 7.2 are assigned corresponding coefficients denoted as $\beta_1, \beta_2, \ldots, \beta_{25}$, the joint utility of living at a location outside the District of Columbia, owning a single family dwelling, owning zero autos, and taking transit to work can be written as

$$
U(\text{live outside D.of C, own single family dwelling, zero autos, transit to work}) = 
\beta_9 \text{TOTIME} + \beta_{10} \text{OVTT/DIST} + \beta_{12} \ln Z + \beta_{15} \text{GPTINV} + 
\beta_{18} \text{HHSIZE1} + \beta_{19} \text{INCDIFF} + \beta_{20} \text{FBFORW} + \beta_{21} \text{FPOB} + 
\beta_{22} \text{DENSITY} + \beta_{23} \text{SCHOOL} + \beta_{25} \ln N_1.
$$

In this case, all the alternative specific constant terms are zero and their coefficients therefore do not appear in the utility function. A somewhat more complex utility function for an alternative characterized by living inside the District in a rented walk-up apartment, owning one auto and using it to commute is as follows:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - DRENT1</td>
<td>[ 1 \text{ in the rent single family dwelling alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>2 - DRENTG</td>
<td>[ 1 \text{ in the rent garden style or walk-up apartment alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>3 - DRENTH</td>
<td>[ 1 \text{ in the rent high rise apartment alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>4 - DA01</td>
<td>[ 1 \text{ in the one auto alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>5 - DA02</td>
<td>[ 1 \text{ in the two or more auto alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>6 - DCAR</td>
<td>[ 1 \text{ in the car to work alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>7 - DAPTSTYL</td>
<td>[ 1 \text{ in the rent garden style, walk-up or high rise apartment and own less than two autos alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>8 - DSUBSTYL</td>
<td>[ 1 \text{ in the own single family dwelling and own two or more autos alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>9 - TOTIME</td>
<td>[ \text{total two way travel time (in min.)} ]</td>
</tr>
<tr>
<td>10 - OVTT/DIST</td>
<td>[ \text{two way out-of-vehicle time (in min.)} \begin{cases} \text{two way travel distance (in miles)} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>11 - DCITY</td>
<td>[ 1 \text{ for households with downtown workplaces in the car to work alternatives} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>12 - ln Z</td>
<td>[ \text{natural logarithm of remaining income} ]</td>
</tr>
<tr>
<td>13 - AALD</td>
<td>[ \frac{\text{number of autos in alternative}}{\text{number of licensed drivers in the household car to work alternatives}} \begin{cases} 1 \text{ otherwise} \ 0 \end{cases} ]</td>
</tr>
<tr>
<td>14 - ILD</td>
<td>[ \frac{1}{\text{number of licensed drivers in the one auto alternatives}} \begin{cases} 0 \text{ for zero auto alternatives} \ 1 \text{ otherwise} \ \frac{2}{\text{number of licensed drivers in the two auto alternatives}} \end{cases} ]</td>
</tr>
</tbody>
</table>

Table 7.2
DEFINITION OF VARIABLES
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 - GPTINV</td>
<td>(1/\text{generalized shopping price by transit})</td>
</tr>
<tr>
<td>16 - R1</td>
<td>(\begin{cases} \text{expected generalized car cost for shopping} \ \text{expected generalized transit cost for shopping} \ 0 \text{ otherwise} \end{cases}) in the one auto alternative</td>
</tr>
<tr>
<td>17 - R2</td>
<td>same as R1 but for two auto alternatives</td>
</tr>
<tr>
<td>18 - HHSIZE1</td>
<td>(\begin{cases} \text{household size in single family dwelling alternatives (own or rent)} \ 0 \text{ otherwise} \end{cases})</td>
</tr>
<tr>
<td>19 - INCDIFF</td>
<td>(\begin{cases} \text{squared income differential when household income exceeds average tract income} \ 0 \text{ otherwise} \end{cases})</td>
</tr>
<tr>
<td>20 - FBFORW</td>
<td>(\begin{cases} \text{fraction non-white households in tract for whites} \ 0 \text{ for non-whites} \end{cases})</td>
</tr>
<tr>
<td>21 - FBFORB</td>
<td>(\begin{cases} \text{fraction of non-white households in tract for non-whites} \ 0 \text{ for whites} \end{cases})</td>
</tr>
<tr>
<td>22 - DENSITY</td>
<td>(\text{net residential density in households per acre})</td>
</tr>
<tr>
<td>23 - SCHOOL</td>
<td>(\begin{cases} \text{per pupil school expenditure for households with children (in dollars per year), except in District of Columbia} \ 0 \text{ in District of Columbia} \end{cases})</td>
</tr>
<tr>
<td>24 - DOC</td>
<td>(\begin{cases} 1 \text{ in District of Columbia} \ 0 \text{ otherwise} \end{cases})</td>
</tr>
<tr>
<td>25 - ln N_i</td>
<td>(\text{natural logarithm of the number of dwelling units in the group of alternatives})</td>
</tr>
</tbody>
</table>

Table 7.2 cont.

DEFINITION OF VARIABLES
U (live in D. of C., rent walk-up apartment, one auto, car to work) =
\[ \beta_2 + \beta_4 + \beta_6 + \beta_7 + \beta_9 \text{TOTIME} + \beta_{10} \text{OVTT/DIST} + \beta_{10} \ln Z + \beta_{13} \text{AALD} \]
\[ + \beta_{14} \text{ILD} + \beta_{15} \text{GPINV} + \beta_{16} \text{R1} + \beta_{19} \text{INCDIFF} + \beta_{20} \text{FBFORW} + \beta_{21} \text{FBFORB} \]
\[ + \beta_{22} \text{DENSITY} + \beta_{24} + \beta_{25} \ln N_i. \]

A final way of viewing the joint model specification is to examine how the value of the variables change across different choice groups. For example, locational variables such as DENSITY change in value only for different locations, and not for different housing, auto ownership, or mode to work alternatives. Other variables may take different values in more than one choice group. For example, AALD, the autos per licensed driver variable, changes for different auto ownership and mode to work options. Furthermore, some variables measure socioeconomic attributes of the household, while others do not.

Table 7.3 summarizes how each variable changes with respect to the four choice groups which comprise the mobility bundle and whether the socioeconomic characteristics of the household (aside from job location) affect the value of each variable. A check appears in those columns for which the variable changes value.

Aside from providing useful insight into the joint model specification, this table allows any conditional probability function to be readily constructed. For example, some of the models to be discussed in this chapter are conditional location and housing models in which auto ownership and mode to work are assumed to be given. All variables in Table
<table>
<thead>
<tr>
<th>Variable</th>
<th>Location</th>
<th>Housing</th>
<th>Auto Ownership</th>
<th>Mode to Work</th>
<th>Socio-economic Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DRENT1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 DRENTG</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 DRENTH</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 DAO1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 DAO2</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 DCAR</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>7 DAPTSTYL</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 DSUBSTYL</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 TOTIME</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>10 OVTT/DIST</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 DCITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 ln Z</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>13 AALD</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>14 ILD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>15 GPTINV</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 R1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>17 R2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>18 HHSIZE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>19 INCDIFF</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 FBFORW</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 FBFORB</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>22 DENSITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 SCHOOL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>24 DOC</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 ln N_i</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3

VARIABLES AND THE CHOICES TO WHICH THEY PERTAIN
7.3 which do not change their value across either the location or housing alternatives do not appear in the model; hence coefficients for the following six variables, DA01, DA02, DCAR, DCITY, AALD, and ILD, are not estimated in the conditional models.

7.4 The Set of Available Alternatives

As discussed in Chapter 5, the underlying theory of the logit model requires that the choice set for each observation consist of feasible alternatives. This implies that in order to properly estimate a joint mobility choice model, one must know which of the enormous set of possible alternatives are actually available to the household. However, questions about the available mobility alternatives were not included in the Washington survey, nor are they likely to be a part of other surveys. Furthermore, if such questions were included, they would have limited usefulness due to various reporting biases (Watson, 1971) and the use in this study of objective or estimated data rather than perceived.

This does not mean that the issue of available alternatives can be ignored. Even without data on the reported choice set, it is possible to state with fairly high reliability that some households will not consider some alternatives as being relevant. For example, households without drivers only have the alternative of zero auto ownership—transit to work. Thus, only location and housing alternatives are
available to them. Another possible restriction of available options is that locations in fringe suburban areas do not have a transit to work option, and hence such locations have associated with them a smaller set of auto ownership/mode to work alternatives.

Using a series of rules such as these, it is possible to eliminate many feasible alternatives. This process has been termed "screening" the alternative set, and is an important part of the modelling process. Failure to do this will result in estimates which are biased and inconsistent, and will therefore produce unreliable forecasts of future conditions. Fortunately, one inherent property of the multinomial logit model is that the screening of relevant (i.e. non-zero probability) alternatives does not result in inconsistent coefficient estimates. Thus, if an error is to be made in screening, it is best to eliminate relevant options from the choice set rather than include irrelevant ones.

This property was used to greatly restrict the set of available location alternatives. Rather than attempt to hypothesize whether or not each tract is actually a feasible alternative for a household with a given workplace, for each work zone a list of the locations actually selected by households in the data sample was constructed. If households with a worker in a given workplace actually chose a residential location, it is apparent that it had a non-zero choice probability to at least those selecting it, and it seems likely that the other households associated with the workplace also perceived the location as a
relevant alternative. Obviously, this technique results in a subset of the entire list of relevant locations, but since the number of alternatives is already quite large the loss of efficiency in the parameter estimation is minimal.

This method of finding relevant locations has the added advantage of selecting a subset of relatively high probability options. Those residential locations actually selected by households are obviously more likely to have high choice probabilities than low ones. Thus, of the relevant locations eliminated by this screening procedure most probably lie at the "tail" of the multinomial logit function, where, since many of the choice probabilities are already quite small, they intuitively provide less information about the shape of the curve than do higher probability options.

Other alternatives were eliminated because certain data items were not available. Rules were defined to eliminate the following alternatives:

(1) Tracts with census codes outside the WCOC cordon area.

(2) Locations in the same tract as the workplace.

(3) Locations for which generalized shopping price data were unavailable.*

(4) Location/housing combinations for which housing cost information was suppressed by the census.

(5) Three tracts in the study area which were improperly processed in the course of building the data base.

*This data was available for only a subset of the possible tract/income group combinations, and the additional cost of creating the remaining data was deemed prohibitive.
In addition, a set of rules reflecting behavioral considerations were used in defining the choice set. Those result in the elimination of the following alternatives:

(1) Alternatives for which the number of automobiles exceeded the number of licensed drivers in the household.

(2) Locations where the fraction of households of the opposite race was less than .01.

(3) Transit to work alternative for locations where the origin walk access for the work trip would exceed one half mile.

(4) Transit to work options for all households working where the destination walk access would exceed one half mile.

(5) Car to work options when the worker has no driver's license.

(6) All alternatives for which the value of Z is less than or equal to zero.

The last of these rules requires some explanation. It reflects the hypothesis that a household must have a certain level of disposable income before it can afford the expenditures of a given mobility bundle. A limitation of this rule is that income may in some cases be an inadequate measure of wealth, which is not included in the data set and may also be important in determining the mobility decision. In spite of this potential difficulty, the rule was used in the screening process. (Note that without this rule, the value of ln Z would be undefined for some alternatives, since the log function does not have values less than or equal to zero in its domain).
Also, these rules collectively create another restriction on the set of admissible observations developed in Chapter 6. The small number of observations which selected an alternative which violates these rules were deleted from the data set to avoid the logical inconsistency of having households choose an alternative which is deemed unavailable. These households probably have characteristics which make them atypical in any case, and the overall model structure for them is probably incorrect. For example, someone who collects antique cars might exhibit extremely high auto ownership, yet his reasons for doing so are not reflected in the model structure. Alternatively, these observations may be the result of coding, reporting, or keypunching errors in the data.

7.5 The Conditional Location/Housing Model for Single Worker Households

As discussed above, the estimation of conditional location and housing models offers a relatively low cost means of estimating most of the joint model coefficients. However, the resulting parameter estimates for the nineteen parameters are substantially less efficient (i.e., have higher variance) than those of the corresponding joint model.

Two different model estimations were performed. The first allowed the coefficient of the variable representing the size of the group of location and housing alternatives, \( \ln N_i \), to attain its maximum likelihood value. These estimates are the first column of figures in Table 7.4. The second set of estimates in this table are based on the constraint that the coefficient of the tract size variable is unity.
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRENTL</td>
<td>-0.212 (-0.615)</td>
<td>0.560 (1.71)</td>
</tr>
<tr>
<td>2</td>
<td>DRENTG</td>
<td>2.08 (2.80)</td>
<td>2.70 (3.56)</td>
</tr>
<tr>
<td>3</td>
<td>DRENTH</td>
<td>0.517 (.686)</td>
<td>0.495 (.637)</td>
</tr>
<tr>
<td>7</td>
<td>DAPTSTYL</td>
<td>0.907 (1.68)</td>
<td>0.905 (1.63)</td>
</tr>
<tr>
<td>8</td>
<td>DSUBSTYL</td>
<td>0.210 (.499)</td>
<td>0.0915 (.216)</td>
</tr>
<tr>
<td>9</td>
<td>TOTIME</td>
<td>-0.00615 (-1.54)</td>
<td>-0.0057 (-1.36)</td>
</tr>
<tr>
<td>10</td>
<td>OVTI/DIST</td>
<td>-0.292 (-0.364)</td>
<td>-0.0260 (-0.312)</td>
</tr>
<tr>
<td>12</td>
<td>ln Z</td>
<td>0.669 (1.78)</td>
<td>0.767 (1.93)</td>
</tr>
<tr>
<td>15</td>
<td>GPTINV</td>
<td>2.53 (1.15)</td>
<td>2.88 (1.28)</td>
</tr>
<tr>
<td>16</td>
<td>R1</td>
<td>-0.922 (-0.707)</td>
<td>-1.14 (-0.854)</td>
</tr>
<tr>
<td>17</td>
<td>R2</td>
<td>-3.13 (-2.24)</td>
<td>-3.21 (-2.26)</td>
</tr>
<tr>
<td>18</td>
<td>HHSTZ51</td>
<td>0.772 (5.29)</td>
<td>0.795 (5.24)</td>
</tr>
<tr>
<td>19</td>
<td>INCDIFF</td>
<td>-0.0106 (-2.64)</td>
<td>-0.0106 (-2.59)</td>
</tr>
<tr>
<td>20</td>
<td>FBFORW</td>
<td>-2.08 (-3.62)</td>
<td>-2.10 (-3.58)</td>
</tr>
<tr>
<td>21</td>
<td>FBFORB</td>
<td>1.91 (2.20)</td>
<td>1.80 (2.08)</td>
</tr>
<tr>
<td>22</td>
<td>DENSITY</td>
<td>-0.00497 (-1.08)</td>
<td>-0.00759 (-1.58)</td>
</tr>
<tr>
<td>23</td>
<td>SCHOOL</td>
<td>0.000424 (.655)</td>
<td>0.000359 (.548)</td>
</tr>
<tr>
<td>24</td>
<td>DOC</td>
<td>-0.0451 (-0.0941)</td>
<td>-0.151 (-0.308)</td>
</tr>
<tr>
<td>25</td>
<td>ln N1</td>
<td>0.487 (5.38)</td>
<td>1 *</td>
</tr>
</tbody>
</table>

| L*(Q) | -690.0 | -690.0 |
| L*(θ) | -583.8 | -597.5 |
| $\chi^2$ | 212.4 | 185.0 |
| NOBS  | 191    | 191    |
| NCASES| 9763   | 9763   |
| percent right | 17.3% | 16.2% |

Table 7.4

ESTIMATES FOR CONDITIONAL LOCATION/HOUSING MODEL: SINGLE WORKER HOUSEHOLDS
For each model, the asymptotic "t" statistics are given parenthesis below their corresponding parameter estimates. In addition, six summary statistics are given, defined as follows:

(1) $L^*(0)$ - the value of the log likelihood function when all parameters are zero (i.e. when every alternative has the same probability);

(2) $L^*(\hat{\beta})$ - the value of the log likelihood function at the maximum likelihood coefficient values.

(3) $\chi^2$ - a statistic equal to $-2(L^*(0) - L^*(\hat{\beta}))$, asymptotically distributed as chi square with the number of degrees of freedom equal to the number of parameters estimated. This statistic provides a test against the null hypothesis that all parameters are zero;

(4) NOBS - the number of households in the sample;

(5) NCASES - the number of available alternatives (in excess of one per household) used in the estimation;

(6) percent right - the percentage of households for which the alternative with the highest systematic component of utility was actually selected. This value is maximized when Manski's maximum score estimation technique is used.

All of the coefficient estimates in both the unconstrained and constrained models for variables about which hypotheses were formulated have the expected sign. However, the statistical significance of some coefficients is quite marginal, particularly for the estimate for OVT/ DIST. This probably results from the very small sample used, since mode choice models with larger samples of the Washington data result in estimates significantly different from zero at fairly high levels of confidence.
The lack of significance of some of the estimates of the coefficients for dummy variables such as DSUBSTYL, the suburban lifestyle constant, is somewhat misleading. The "t" statistic can change depending on the alternative used as a reference in the definition of the alternative specific constants. Only the standard error of the estimate of the difference between the utilities in general remains invariant with respect to the base used.

One potential area for concern in these results is the fact that the unconstrained estimate of the coefficient of ln N_i is statistically different from unity at a very high confidence level. This might be attributed to a number of factors, including some systematic measurement error obtained when converting census counts to match the definition of housing type in the home interview survey. It may also result from a failure to some degree of the independence axiom implicit in the grouping of alternatives. Most likely it results from a combination of both of these factors as well as some possible misspecification of the utility function.

Most of the shift in the coefficient estimates caused by the use of constrained estimation is in the value of constant terms. (The value for DRENTL, the constant for renting single family housing, actually reverses sign). This offers some limited support for attributing the small value for the coefficient of the grouped alternative variable, ln N_i, to
measurement error, since the constant terms can compensate for such errors by shifting their value. Complete failure of the independence axiom or some major model mis-specification would seem far more likely to lead to a general instability in all the coefficient estimates.

Both sets of estimates, when taken as a whole, are statistically significantly different from zero. However, this test is obviously quite weak, since the null hypotheses of equally likely alternatives is somewhat unreasonable.* As expected, the use of the constrained estimation reduces the "goodness of fit" of the model as measured by the value of the log likelihood function.

7.6 The Joint Mobility Choice Model for Single Worker Households

While the conditional model described in the previous section offers a low cost method of examining a subset of the coefficients of the joint model, the most efficient estimate of all the coefficients is generally obtained with joint estimation of the entire model. The unconstrained and constrained estimates for this model are presented in Table 7.5. Note that the number of observations used in estimating the joint model is somewhat less than in the conditional estimation. The eliminated observations resulted in lists of variables exceeding computer limits, and were therefore eliminated to avoid extra data processing.

The parameter estimates for all variables for which strong hypotheses were established have the expected sign. Furthermore, as anticipated, the standard errors of the estimates of coefficients appearing in both the conditional and joint models are reduced.

*A failure to pass this test would probably indicate mis-specification of major importance.
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRENT1</td>
<td>-0.361 (.103)</td>
<td>0.393 (1.18)</td>
</tr>
<tr>
<td>2</td>
<td>DRENTG</td>
<td>2.31 (2.87)</td>
<td>2.93 (3.58)</td>
</tr>
<tr>
<td>3</td>
<td>DRENTH</td>
<td>0.828 (.102)</td>
<td>0.809 (.973)</td>
</tr>
<tr>
<td>4</td>
<td>DAO1</td>
<td>7.86 (2.57)</td>
<td>7.98 (2.60)</td>
</tr>
<tr>
<td>5</td>
<td>DAO2</td>
<td>12.0 (2.71)</td>
<td>12.1 (2.78)</td>
</tr>
<tr>
<td>6</td>
<td>DCAR</td>
<td>0.433 (.500)</td>
<td>0.483 (.501)</td>
</tr>
<tr>
<td>7</td>
<td>DAPTSTYL</td>
<td>0.342 (.966)</td>
<td>0.524 (.927)</td>
</tr>
<tr>
<td>8</td>
<td>DSUBSTYL</td>
<td>0.336 (.764)</td>
<td>0.261 (.591)</td>
</tr>
<tr>
<td>9</td>
<td>TOTIME</td>
<td>-0.00831 (-2.13)</td>
<td>-0.00818 (-2.05)</td>
</tr>
<tr>
<td>10</td>
<td>OVIT/DIST</td>
<td>-0.0570 (-.787)</td>
<td>-0.0526 (-.708)</td>
</tr>
<tr>
<td>11</td>
<td>DCITY</td>
<td>-0.437 (-.932)</td>
<td>-0.415 (-.879)</td>
</tr>
<tr>
<td>12</td>
<td>ln Z</td>
<td>1.07 (2.64)</td>
<td>1.20 (2.81)</td>
</tr>
<tr>
<td>13</td>
<td>AALD</td>
<td>0.964 (1.01)</td>
<td>0.975 (1.02)</td>
</tr>
<tr>
<td>14</td>
<td>ILD</td>
<td>-6.57 (-2.17)</td>
<td>-6.56 (-2.16)</td>
</tr>
<tr>
<td>15</td>
<td>GPTINV</td>
<td>2.92 (1.38)</td>
<td>3.14 (1.47)</td>
</tr>
<tr>
<td>16</td>
<td>R1</td>
<td>-1.33 (-1.08)</td>
<td>-1.34 (-1.21)</td>
</tr>
<tr>
<td>17</td>
<td>R2</td>
<td>-4.05 (-3.01)</td>
<td>-4.11 (-3.03)</td>
</tr>
<tr>
<td>18</td>
<td>HHSIZE1</td>
<td>0.850 (5.21)</td>
<td>0.875 (5.16)</td>
</tr>
<tr>
<td>19</td>
<td>INCDIFF</td>
<td>-0.0123 (-2.89)</td>
<td>-0.0121 (-2.80)</td>
</tr>
<tr>
<td>20</td>
<td>FBFORW</td>
<td>-2.18 (-3.79)</td>
<td>-2.21 (-3.78)</td>
</tr>
<tr>
<td>21</td>
<td>FBFORB</td>
<td>1.95 (2.23)</td>
<td>1.85 (2.12)</td>
</tr>
</tbody>
</table>

Table 7.5
ESTIMATES FOR JOINT MOBILITY MODEL: SINGLE WORKER HOUSEHOLDS
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>DENSITY</td>
<td>-.00557 (-1.25)</td>
<td>-.00810 (-1.75)</td>
</tr>
<tr>
<td>23</td>
<td>SCHOOL</td>
<td>.000442 (.685)</td>
<td>.000342 (.523)</td>
</tr>
<tr>
<td>24</td>
<td>DOC</td>
<td>-.00993 (-2.06)</td>
<td>-.100 (-2.04)</td>
</tr>
<tr>
<td>25</td>
<td>ln N_i</td>
<td>.492 (5.25)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>L*(Q)</td>
<td>-824.4 (5.25)</td>
<td>-824.4</td>
</tr>
<tr>
<td></td>
<td>L*(β)</td>
<td>-645.9 357.0</td>
<td>-658.4 332.0</td>
</tr>
<tr>
<td></td>
<td>χ²</td>
<td>177                  177</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NOBS</td>
<td>25601                25601</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NCASES</td>
<td>25601                25601</td>
<td></td>
</tr>
<tr>
<td></td>
<td>percent right</td>
<td>8.5%</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

*constraint imposed, hence "t" statistic not relevant

Table 7.5 cont.
ESTIMATES FOR JOINT MOBILITY MODEL: SINGLE WORKER HOUSEHOLDS
It is interesting to compare the estimates of the joint and conditional models. The conditional estimates all lie within a standard error (taken from the joint model) of the joint estimates, and most are much closer. This high degree of stability provides some confirmation that the use of a joint estimation for mobility decisions does not produce disturbance terms of the utility function which are greatly different from those of the conditional model.*

The constrained estimates again are quite similar to the unconstrained ones, with the exception of the coefficient of DRENTL. This suggests the possibility of some measurement error in the value of $N_i$, the number of units of type $i$ in a tract, for rented single family dwellings, though this could not be confirmed upon examination of the census data processing programs.

The overall goodness of fit of both sets of estimates indicates a highly significant improvement over the hypothesis of zero parameters. As one might expect when an average of over 145 alternatives are available for each household, the percent right statistic is quite low in absolute terms. However, a useful way of viewing this statistic is in terms of the probability of a model classifying a given percent correctly if all alternatives were actually equally likely.

Suppose that all households had exactly 145 available alternatives, and that each was equally likely. In this case, the probability of a model classifying none of the 177 observations correctly is simply

---

*McFadden suggests this possibility as a reason for avoiding joint model estimation.
\[
\frac{(144)^{177}}{(145)} = .2938
\]

More generally, the probability of classifying \( k \) of 177 correctly is distributed as binomial, i.e.

\[
Pr \text{ (k correct)} = \binom{177}{k} (1/145)^k (144/145)^{177-k}.
\]

Using this formula, the odds of classifying 9 or more correct (about 5% right) is less than .0001, yet the percent right for the unconstrained and constrained estimates are 8.5% and 10.2% respectively.*

7.7 The Conditional Location/Housing Model for Multi-Worker Households

The model specified in Section 7.3 is structured to represent the behavior of single worker households. As discussed in the theoretical development in Chapter 4, the mobility decision process for multi-worker households is far more complex, since there is more than one work trip involved in the location decision. This important behavioral consideration leads to the estimation of a separate model for multi-worker households.

The four causal mechanisms proposed in Chapter 4 for the location decision of multi-worker households correspond to the following behavioral hypotheses:

*It is interesting to note that the percent right is actually higher for the constrained estimates than for the unconstrained, indicating that as expected Manski's maximum score estimates differ from those obtained by the maximum likelihood method.
(1) complete primary worker dominance;

(2) a primary worker with the remaining workers secondary (i.e. without fixed workplaces) in the location decision;

(3) some or all workers with fixed workplaces but each with different weights;

(4) complete equality in the perception of the work trip attributes.

Each of these hypotheses can at least in theory be represented in a functional form of the utility function; however, each resulting form has different data and computational requirements which may affect the model's ultimate usefulness.

The assumption of one completely dominant worker implies that the model specification for multi-worker households is identical to that estimated for single worker families. Under this hypothesis, there exists a primary worker, whose identity would obviously have to be determined \textit{a priori}. Only this worker's level of service would appear in the utility function, and only his mode to work would be part of the mobility bundle. Note that it still might be necessary to estimate separate mobility choice models for single and multiple worker households, since the fact that other members of the household work might well alter the relative importance of shopping accessibility, the way in which autos are used in the household (and hence the desirability of auto ownership) and even the importance of various location and housing attributes to the household.
A model reflecting this extremely naive hypothesis was estimated. As one might anticipate, it yielded unreasonable results in both the unconstrained and constrained forms. In particular, the coefficients of the level of service measures, TOTIME and OVTT/DIST, were positive and had fairly high standard errors.

This anticipated failure lead to the formulation of measures of accessibility to employment to reflect the second possible causal mechanism. Under this hypothesis, only the primary worker's mode and level of service attributes are part of the mobility bundle, but the overall desirability of a location depends on its proximity to other places of employment.

Ideally, the measures of employment accessibility used would be directly analogous to the shopping accessibility measures; however, this would require an employment location choice model which might be as difficult to develop as the mobility choice models themselves. The alternative employment accessibility measure used is of the isochron type, and resulted in two variables defined as follows:

\[ HIACC_i = \text{fraction of regional employment which is within 45 minutes of tract } i \text{ by car in the peak period}; \]

\[ TRACC_i = \text{fraction of regional employment which is within 45 minutes of tract } i \text{ by transit in the peak period}. \]

These measures, while rather crude, do provide at least some information about the accessibility to employment. However, even when they
were introduced into the utility function in a variety of ways, the coefficients of the primary worker's level of service remained positive.*

These results lead directly to the conclusion that it may be essential to consider the workplaces of some or all workers in the household as fixed in the mobility decision, and that as a consequence it may prove necessary to explicitly represent every worker's trip attributes in the utility function. This type of model requires an enormous computational effort, since the level of service from every tract to every possible workplace may potentially be used in the estimation.** Furthermore, to use such a model to make forecasts, one must then know within each household the joint location of every worker's workplace rather than simply the workplace of the primary worker.

Because of limitations on the resources available for this study, a somewhat simpler, albeit less satisfactory, model was estimated. Rather than use each distinct level of service attribute as a separate variable, the total of all trip lengths for work trips was used. This variable is defined as follows:

* In general, the coefficient of HIACC had an unexpected sign, perhaps for the same reason that highway generalized shopping price was eliminated from the single worker models. The remaining variable, TRACC, was used in a variety of forms, including ones which reflected both auto ownership and the mode of work for the primary worker.

**Recall that only 58 workplaces were used in this study for the single worker households.
TOTDIST = sum of all one way travel distances for work trips (miles).

TOTDIST varies only across alternative locations; hence, it does not distinguish between the modes used or autos owned. However, it is a crude measure of the desirability of the household's total work trip pattern if it chose a given location as a place of residence.

Table 7.6 presents the results of the estimation of the conditional location/housing choice model for multi-worker households. Note that the level of service measures, TOTIME and OWTT/DIST, do not appear in the model and that the variable TOTDIST is labelled as number 26 to maintain consistency with the convention in Table 7.2.

As with the single worker household models, all coefficients have the expected sign, and the constrained estimates are fairly close to the unconstrained ones. As in the previous conditional model, the statistical significance of many of the estimates is quite low, a problem exacerbated by the even smaller sample of households used. Because of this high unreliability in many parameter estimates, it is difficult to draw inferences from a comparison between the conditional models for single and multi-worker households. For this reason, these comparisons, described in Chapter 8, will only be made between the two joint model forms.

7.8 The Joint Mobility Choice Model for Multi-Worker Households

The specification of the joint mobility choice model for multi-worker households is a direct extension of the models described in the
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRENTL</td>
<td>-0.876 (-1.56)</td>
<td>-0.576 (-1.04)</td>
</tr>
<tr>
<td>2</td>
<td>DRENTG</td>
<td>2.70 (2.50)</td>
<td>2.96 (2.69)</td>
</tr>
<tr>
<td>3</td>
<td>DRENTH</td>
<td>1.28 (1.17)</td>
<td>1.21 (1.09)</td>
</tr>
<tr>
<td>7</td>
<td>DAPTSTYL</td>
<td>1.11 (1.42)</td>
<td>1.25 (1.56)</td>
</tr>
<tr>
<td>8</td>
<td>DSUBSTYL</td>
<td>0.120 (.108)</td>
<td>-0.0443 (-.0649)</td>
</tr>
<tr>
<td>12</td>
<td>ln Z</td>
<td>1.21 (1.90)</td>
<td>1.37 (2.11)</td>
</tr>
<tr>
<td>15</td>
<td>GPTINV</td>
<td>1.68 (.631)</td>
<td>2.16 (.793)</td>
</tr>
<tr>
<td>16</td>
<td>R1</td>
<td>-1.08 (-.588)</td>
<td>-1.27 (-.677)</td>
</tr>
<tr>
<td>17</td>
<td>R2</td>
<td>-1.32 (-.784)</td>
<td>-1.51 (-.875)</td>
</tr>
<tr>
<td>18</td>
<td>HHSIZE1</td>
<td>1.17 (4.27)</td>
<td>1.24 (4.18)</td>
</tr>
<tr>
<td>19</td>
<td>INCDIFF</td>
<td>-0.00412 (-1.08)</td>
<td>-0.00372 (-.959)</td>
</tr>
<tr>
<td>20</td>
<td>FBFORW</td>
<td>-0.432 (-.563)</td>
<td>-0.444 (-.560)</td>
</tr>
<tr>
<td>21</td>
<td>FBFORB</td>
<td>3.23 (3.04)</td>
<td>3.30 (3.02)</td>
</tr>
</tbody>
</table>

Table 7.6
ESTIMATES FOR CONDITIONAL LOCATION/HOUSING MODEL:
MULTI-WORKER HOUSEHOLDS
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>DENSITY</td>
<td>$-0.00740$ ($-0.727$)</td>
<td>$-0.00889$ ($-0.867$)</td>
</tr>
<tr>
<td>23</td>
<td>SCHOOL</td>
<td>$0.000279$ ($0.300$)</td>
<td>$0.000204$ ($0.213$)</td>
</tr>
<tr>
<td>24</td>
<td>DOC</td>
<td>$-1.04$ ($-1.43$)</td>
<td>$-1.18$ ($-1.58$)</td>
</tr>
<tr>
<td>25</td>
<td>ln $N_1$</td>
<td>$0.703$ ($4.78$)</td>
<td>$1$ *</td>
</tr>
<tr>
<td>26</td>
<td>TOTDIST</td>
<td>$-0.00106$ ($-1.19$)</td>
<td>$-0.00103$ ($-1.14$)</td>
</tr>
<tr>
<td></td>
<td>$L^*(0)$</td>
<td>$-347.9$</td>
<td>$-347.9$</td>
</tr>
<tr>
<td></td>
<td>$L^*(\hat{\beta})$</td>
<td>$-277.2$</td>
<td>$-279.0$</td>
</tr>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>$141.4$</td>
<td>$137.8$</td>
</tr>
<tr>
<td></td>
<td>NOBS</td>
<td>$105$</td>
<td>$105$</td>
</tr>
<tr>
<td></td>
<td>NCASES</td>
<td>$4231$</td>
<td>$4231$</td>
</tr>
<tr>
<td></td>
<td>percent right</td>
<td>$20.0%$</td>
<td>$21.9%$</td>
</tr>
</tbody>
</table>

*constraint imposed, hence "t" statistic not relevant

Table 7.6 (cont.)
ESTIMATES FOR CONDITIONAL LOCATION/HOUSING:
MULTI-WORKER HOUSEHOLDS
As in the conditional model for these households, the variables TOTIME and OVT/DIST describing the level of service for work trips are omitted, and TOTDIST, the total work trip distance summed over all workers, is included. The elimination of the travel time variables not only reduces the level of sophistication in the model's description of location choice, but also reduces the descriptive power of the mode to work decision. In this model, choice of mode for the primary worker is a function of a constant term, travel cost, auto ownership and the primary worker's place of employment (whether or not it is in the CBD). Furthermore, only the mode selected by the primary worker is an explicit choice in the model. Thus, the resulting specification is a compromise between the use of a total primary worker approach and one which doesn't distinguish among workers in the household.

The estimation results for this model are presented in Table 7.7. As with the single worker household models, for computational reasons the joint model sample size is somewhat less than in the corresponding conditional model. Perhaps the most notable change in the parameters between the conditional and joint models is the drop in magnitude of the value of the coefficient of TOTDIST, resulting in a decrease in its statistical significance.* This drop, while within a standard error of the conditional estimates, seems quite large as compared with the corresponding changes in the single worker household models, and may result from the mis-specification in the description of the primary worker's mode choice decision.

* The standard error in the joint model only increased marginally.
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrained Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRENT1</td>
<td>-1.21 (-1.97)</td>
<td>-.924 (-1.52)</td>
</tr>
<tr>
<td>2</td>
<td>DRENTG</td>
<td>3.14 (2.53)</td>
<td>3.45 (2.78)</td>
</tr>
<tr>
<td>3</td>
<td>DRENTH</td>
<td>1.47 (1.18)</td>
<td>1.44 (1.15)</td>
</tr>
<tr>
<td>4</td>
<td>DA01</td>
<td>2.25 (1.13)</td>
<td>2.38 (1.18)</td>
</tr>
<tr>
<td>5</td>
<td>DA02</td>
<td>4.79 (1.80)</td>
<td>5.08 (1.90)</td>
</tr>
<tr>
<td>6</td>
<td>DCAR</td>
<td>2.51 (1.89)</td>
<td>2.51 (1.89)</td>
</tr>
<tr>
<td>7</td>
<td>DAPTSTYL</td>
<td>1.21 (1.45)</td>
<td>1.30 (1.55)</td>
</tr>
<tr>
<td>8</td>
<td>DSUBSTYL</td>
<td>.0168 (.0226)</td>
<td>-.145 (-.194)</td>
</tr>
<tr>
<td>11</td>
<td>DCITY</td>
<td>-2.34 (-2.70)</td>
<td>-2.33 -2.69</td>
</tr>
<tr>
<td>12</td>
<td>ln Z</td>
<td>1.71 (2.46)</td>
<td>1.87 (2.66)</td>
</tr>
<tr>
<td>13</td>
<td>AALD</td>
<td>1.05 (.727)</td>
<td>1.04 (.723)</td>
</tr>
<tr>
<td>14</td>
<td>ILD</td>
<td>-1.86 (-.887)</td>
<td>-1.85 (-.880)</td>
</tr>
<tr>
<td>15</td>
<td>GPTINV</td>
<td>1.10 (-3.92)</td>
<td>1.57 (.549)</td>
</tr>
</tbody>
</table>

Table 7.7
ESTIMATES FOR JOINT MOBILITY MODEL: MULTI-WORKER HOUSEHOLDS
<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unconstrained Estimates</th>
<th>Constrasted Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>R1</td>
<td>-.418 (-2.22)</td>
<td>-.616 (-.321)</td>
</tr>
<tr>
<td>17</td>
<td>R2</td>
<td>-1.97 (-1.13)</td>
<td>-2.19 (-1.23)</td>
</tr>
<tr>
<td>18</td>
<td>HHSIZE1</td>
<td>1.42 (4.25)</td>
<td>1.49 (4.42)</td>
</tr>
<tr>
<td>19</td>
<td>INCDIFF</td>
<td>-.00511 (-1.17)</td>
<td>-.00471 (-1.07)</td>
</tr>
<tr>
<td>20</td>
<td>FBFORW</td>
<td>-.797 (-1.02)</td>
<td>-.822 (-1.04)</td>
</tr>
<tr>
<td>21</td>
<td>FBFORB</td>
<td>2.90 (2.67)</td>
<td>2.94 (2.68)</td>
</tr>
<tr>
<td>22</td>
<td>DENSITY</td>
<td>-.00775 (-.697)</td>
<td>-.00917 (-.833)</td>
</tr>
<tr>
<td>23</td>
<td>SCHOOL</td>
<td>.000433 (.438)</td>
<td>.000353 (.355)</td>
</tr>
<tr>
<td>24</td>
<td>DOC</td>
<td>-.609 (-.831)</td>
<td>-.731 (-1.00)</td>
</tr>
<tr>
<td>25</td>
<td>ln Ni</td>
<td>.705 (4.41)</td>
<td>1 *</td>
</tr>
<tr>
<td>26</td>
<td>TOTDIST</td>
<td>-.000324 (-.312)</td>
<td>-.000244 (-.237)</td>
</tr>
</tbody>
</table>

\[ L^*(0) = -422.3 \]
\[ L^*(\hat{\beta}) = -297.4 \]
\[ \chi^2 = 249.8 \]
\[ NOBS = 95 \]
\[ NCASES = 11671 \]
\[ percent \ right = 20.0\% \]

Table 7.7 (cont.)
ESTIMATES FOR JOINT MOBILITY MODEL: MULTI-WORKER HOUSEHOLDS
As in the previous estimations, the use of the constrained tract size variable did not significantly affect the remaining coefficients. Furthermore, as in all previous results, the estimated coefficients, when taken as a whole, are significantly different from zero at extremely high confidence levels.

Many of the estimated coefficients have extremely small values and are not statistically different from zero at even very low confidence levels. Some variables such as DSUBSTYL might have been eliminated in later estimations except for the desire to maintain comparability between the single and multi-worker household models.

7.9 Summary of Estimation Results

This chapter has developed specifications of a joint mobility choice model for both single and multi-worker households. These specifications reflect a broad range of causal factors, including measures in each of the six categories of variables described in Chapter 2. Even using the relatively small samples required due to budget limitations, the results indicate that it is possible to consider an extensive number of causal hypotheses using a joint disaggregate choice model and that this approach, when applied in the context of a larger study, is likely to prove extremely fruitful.

A subset of the parameters of these models was estimated using the conditional location and housing decision, given automobile ownership and
mode to work. These estimates, while consistent and less expensive to obtain than those from the joint model, tend to have relatively high variance, since they utilize only a small fraction of the total data available. Estimates for the joint models were also developed, and a significant improvement in the estimates was obtained even when a smaller number of observations was used.

In general, the single worker household models are superior to their multi-worker counterparts. This results from the difficulty inherent in representing the work trip attributes of all the workers jointly. Because of the limitations on available computer time, the final multi-worker models are a compromise between the computational ease of using only the primary worker's workplace as fixed and the apparent true behavioral mechanism in which the household considers the fixed workplaces of more than one worker when selecting a mobility bundle.

The amount of computation time required to estimate these models varied substantially, depending on the size of the data set, the number of iterations needed for the estimation program to converge and the number of coefficients estimated.* (Central processing charges are by far the most significant computation cost item.) For the joint single worker and multi-worker household models, the computation time required per iteration on an IBM 370/168 was approximately .71 and .32 minutes respectively. The unconstrained and constrained estimations had virtually identical computational requirements.

*The estimation program uses the Newton-Raphson method for solving a system of non-linear equations. This technique is an iterative procedure requiring the solution of a series of linear approximations.
All but one of the joint model estimation runs required four iterations to converge to relatively tight tolerance limits. The remaining run required six iterations.

In addition to these costs, the equivalent of roughly one iteration's central processing time was required to set up the input data for the estimation program.

The next chapter is a discussion of some of the behavioral implications of the model results. Most of the inferences are drawn from an analysis of the single worker models, though some attempt is made to compare aspects of the behavior of single and multi-worker families.
Chapter 8

BEHAVIORAL IMPLICATIONS OF THE MODELS

8.1 Evaluating the Implications of the Models

The models presented in the preceding chapter represent the effect of a broad spectrum of causal factors on the mobility decisions of urban households. These effects are sufficiently complex in nature that merely examining the coefficient estimates of the utility functions provides little insight into the behavioral implications of the results. This chapter is an attempt to use the models described in Chapter 7 to draw behavioral inferences about household mobility behavior, and, in at least a preliminary way, to assess the possible impact of a number of transportation policies.

Obtaining useful and unambiguous insights from a complex, joint disaggregate model is not as straightforward as one might think. First, the units of the utility function are difficult to interpret. Merely stating that the utility of one alternative is some quantity greater than that of another is in itself of very little value. An alternative approach is to consider the choice probabilities implied by the utility values. This measure at least has the advantage that given a sufficient number of identical decision-makers, the choice probability of an alternative corresponds to its expected share in the group. However, if the number of alternatives is quite large it may be too cumbersome to consider all the choice probabilities.
A second problem associated with examining a complex model is that it is frequently meaningless to separate out various causal effects. In the real world, independent variables often tend to move together. For example, neighborhood quality and housing prices are generally positively correlated, and the level of auto ownership and the expenditures associated with that level are (at least under the assumptions of the model) perfectly correlated. Merely examining the effect of changes in one variable without considering possible covariation can potentially lead to conclusions which are of little relevance to real situations.

The problem of drawing inferences from the model results is further compounded by the randomness inherent in the parameter estimates. It is often tempting to analyze simple point estimates without considering their variance. Furthermore, when an inference is based on some function of the parameter estimates, the variance of the function is often far greater than the variance of any of the individual parameter estimates. This problem will be considered in greater detail later in this chapter.

In order to avoid some of the above difficulties, this chapter relies on a number of approaches to considering the model results. No single summary statistic or evaluation methodology is in itself likely to prove sufficient to either confirm or reject any behavioral hypothesis. Furthermore, the reader is cautioned that all inferences drawn in this chapter should be viewed as preliminary in nature, and as pertaining to one data
set from one city and from one point in time. The ultimate verification of any conclusions must come from a body of empirical research, not a single study. Given this caveat, four basic approaches to evaluating the model results are considered.

First, the coefficients of the single and multi-worker household models are compared. Many of the significant variations between these groups can be attributed to basic behavioral differences between the two household types; others are more difficult to explain, and are possible indications of some subtle mis-specifications in the model or lead to new hypotheses about mobility choice which should ultimately be explored. Following this, the "pure alternative" effects are considered. This latter analysis focuses on the relative desirability of alternatives given that all their other attributes are identical. The third approach used is to examine the "value of time," i.e. the marginal rate of substitution of time and cost implied by the model. Since travel time measures do not appear in the multi-worker households models, this discussion is limited to the behavior of single worker households.

The final evaluation technique used is the prediction of the choice probabilities for a number of "prototypical" households when confronted with a set of mobility choice alternatives. A hypothetical eleven zone city with a CBD workplace was created for this analysis in order to explore the forecast choice probabilities under a limited but relatively realistic choice situation. Using the mobility model for single worker households, the potential impacts of a variety of possible transportation policies are considered.
Two other possible evaluation methods were not used. In order to avoid a major computational problem, no attempt was made to consider the aggregate behavior of a random group of households in the Washington region. Such an analysis of aggregate response, while potentially interesting, was simply infeasible within the resources available to the study.

The other omitted analysis method is the calculation of elasticities. In aggregate models, these measures provide a useful summary of the total percentage response to a particular change in an independent variable. However, in a disaggregate model, the elasticity is generally measured with respect to the choice probability.* In the case of the joint mobility decision, the elasticity would be taken with respect to the probability of a particular location, housing, auto ownership and mode to work combination. It is difficult to determine precisely what such a measure means and what the behavioral implications of its value are. Furthermore, in the multinomial logit model, the elasticities are a function of the choice probabilities and the values of the independent variables, and any single numerical elasticity is extremely case specific. For these reasons, the computation of elasticities was deemed to be the least desirable method of analyzing the model results.

8.2 Comparison of Coefficients

The principal motivation for estimating separate mobility models for single and multi-worker households was the a priori expectation that there

*It is possible to compute aggregate elasticity of a group as a function of every individual’s disaggregate elasticity, but this would require a computational effort equivalent to that of making aggregate forecasts.
are important behavioral differences between the two groups which go beyond their perceptions of the attributes of the work trip. The substantial differences in the coefficient estimates for the joint models in Tables 7.5 and 7.7 tend to confirm this hypothesis for some coefficients, though limitations in the available computer budget prohibited an extensive series of statistical tests to determine which coefficients can be constrained to be equal across groups.* On the other hand, other coefficients seem extremely stable considering their inherent randomness.

The three coefficients which seem the most stable between the two household groups are AALD, DENSITY and SCHOOL, the auto availability, net residential density and per pupil school expenditures respectively. In the constrained models these coefficients have the following values:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Single Worker Estimate</th>
<th>Multi-Worker Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AALD</td>
<td>.975</td>
<td>1.04</td>
</tr>
<tr>
<td>DENSITY</td>
<td>-.00810</td>
<td>-.00917</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>.000342</td>
<td>.000353</td>
</tr>
</tbody>
</table>

Given the fairly high standard errors of most of these coefficients, it is likely that they could be constrained across household groups without significantly altering the other coefficients. Furthermore, the resulting set of constrained estimates would probably be more efficient than

*This test is a maximum likelihood equivalent to the Chow test in least squares estimation, and relies on the theorem that given q linear constraints between two models, -2 times the natural logarithm of the ratio of the constrained and unconstrained likelihoods is asymptotically distributed as chi-square with q degrees of freedom. (Theil, 1971).
either of the two unconstrained sets. The closeness of these estimates implies that single and multi-worker households do not appear to exhibit significant differences in the value they place on low density and school expenditures, nor does auto availability have a differential effect on the preferences for use of the car mode by the primary worker.

Coefficient estimates of some of the other variables differ by a factor of two or more. These variables, a brief mnemonic definition, and the estimated coefficients for both the constrained joint models are presented in Table 8.1.

The difference in the coefficient of ILD, which reflects the increasing preference for auto ownership with increasing numbers of drivers, will be explored further in the following section, since this variable along with the net pure auto ownership effects, DA01 and DA02, and the auto availability variable, AALD, determine the magnitude of the total pure auto ownership effect. Greater values of the magnitude of the coefficient of ILD imply a greater sensitivity of auto ownership to the number of drivers in a household. Thus, an additional driver seems to have a much greater effect on single worker households than on their multi-worker counterparts.

One possible explanation for this is based on the hypothesis that, in general, multi-worker households have the opportunity to share a ride to work; hence, one car can frequently be used to serve more than one work trip even when every worker has a license. By contrast, if
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ILD</td>
<td>inverse licensed drivers</td>
<td>-6.56</td>
<td>-1.85</td>
</tr>
<tr>
<td>GPTINV</td>
<td>inverse of generalized transit shopping price</td>
<td>3.14</td>
<td>1.57</td>
</tr>
<tr>
<td>R1</td>
<td>ratio of car and transit generalized shopping price (for one auto)</td>
<td>-1.54</td>
<td>- .616</td>
</tr>
<tr>
<td>R2</td>
<td>same as R1, but for 2+ autos</td>
<td>-4.11</td>
<td>-2.19</td>
</tr>
<tr>
<td>INCDIFF</td>
<td>squared income differential in tracts w/ lower income</td>
<td>- .0121</td>
<td>- .00471</td>
</tr>
<tr>
<td>FBFORW</td>
<td>fraction black for whites</td>
<td>-2.11</td>
<td>- .882</td>
</tr>
</tbody>
</table>

Table 8.1

COEFFICIENT ESTIMATES FOR PARAMETERS WITH VARIATION ACROSS HOUSEHOLD TYPES
there is only one worker in a household but more than one licensed driver, the daily use of an automobile by one driver most often denies its use to the other. Thus, across a sample of households, additional drivers will overall tend to have greater effect on the auto ownership of single worker families than on those with more than one worker.

The recent development of modal choice models (Cambridge Systematics, 1974) which explicitly include the shared ride alternative lend some support to the above explanation. In these models, travellers from households with large numbers of workers have a significantly higher probability of sharing a ride than those from single worker households. Intra-household shared rides would therefore appear to be a significant possibility for multi-worker households.

In order to empirically test the above explanation it would be necessary to develop a much more sophisticated model of the mode to work aspects of the mobility choice of multi-worker households than the one described previously. This model would have to explicitly represent the possibility of workers within a household combining their work trips, and incorporate the effect of such a decision on auto ownership in a behavioral way.

The next three variables in Table 8.1, GPTINV, R1, and R2, all reflect the accessibility of a given location to shopping opportunities. Two effects are explicitly accounted for in these variables, the
absolute level of transit accessibility and the relative levels of car and transit accessibility. In both these cases, it appears that the single worker households are far more sensitive to the level of accessibility. This can be ascribed to the essential differences in the structure of single and multiple worker households. The former frequently consists of a worker, a spouse responsible for day to day shopping, and children. In such households, shopping is a much more significant function than households in which both adults work, and shopping trips are likely to be made far more frequently than in multi-worker households. For this reason, one might anticipate that the mobility decisions of single worker households are more affected by the level of shopping accessibility a location offers.

The remaining two variables in Table 8.1, INCDIFF and FBFORW, are attributes of locations. As with the spatial opportunity variables, the magnitude of the coefficients for single worker households is greater than for multi-worker families. This may again reflect differences in the life style of the two household types. Multi-worker households are probably far less concerned about neighborhood attributes because working absorbs a large portion of their daily lives. In single worker households, one adult is often at home the entire day, and the household is therefore more interested in establishing social ties in the area in which it lives.
This effect does not seem to be consistent for non-white households, since the corresponding coefficient of the racial composition variable is greater for multi-worker households than for their single worker counterparts.

The remaining coefficients all have values which differ in magnitude but by too little to merit attempts to draw any inferences about the household mobility behavior.

8.3 Pure Alternative Effects

The constant terms in the utility functions measure all of the attributes of an alternative not explicitly represented in the remaining variables. These attributes, when coupled with alternative specific socioeconomic variables, permit the examination of household preferences under rigid ceterus paribus conditions.

Suppose two mobility bundles had exactly identical attributes, except one required the use of the car mode while the other used transit to work.* The pure alternative difference in utility in the constrained single worker family model would then be

\[\text{.483} - \text{.415 DCITY},\]

where the second term is used only for households with a CBD worker. Thus, car is the preferred mode, though only a minor preference is indicated when the worker is employed in the CBD. If the two alternatives

*This is only a sufficient condition; the necessary condition is that the utilities of the bundles (except for mode constants) are equal.
in this example were the only ones available, this difference translates to choice probabilities for the car-related bundle of .618 and .517 for non-CBD and CBD workers respectively.

In multi-worker households, estimates of pure mode choice effects are probably very biased since travel times are omitted from the model. As one might expect, the car bias is much greater, since car is typically the much faster mode. The equivalent choice probabilities for multi-worker households are .925 and .527 respectively.

The pure auto ownership effects are somewhat more complex, since they involve comparisons among three alternatives and are measured by more variables. In the joint model, zero auto ownership is always the base, and thus has zero pure alternative utility. The variables, DA01, DA02, ILD, and, when the car mode is used, AALD, all affect the pure alternative utilities.

Figure 8.1 depicts the pure alternative utility from the constrained estimates of one and two auto ownership as a function of the number of licensed drivers. The pure alternative effect is measured by a "utility score," i.e. the difference between the given alternative and that of owning no autos. No point is plotted for alternatives with a greater number of autos than drivers, since these options are by definition unavailable. There are two sets of lines in this figure; the relevant one depends on the mode to work.
2 cars; single worker households

1 car; single worker households

2 cars; multi-worker households

1 car; multi-worker households

---

Figure 8.1

PURE AUTO OWNERSHIP UTILITIES
When the utility values in this figure are translated into conditional probabilities, the results indicate extremely strong auto ownership preferences. In general, when there is more than one driver in the household the probability of zero auto ownership is below .01 and the probability of owning one auto is almost always less than 2. These results, however, reveal the essential weakness of the \textit{ceterus paribus} reasoning implicit in the analysis of pure auto ownership preferences. Obviously, \textit{all else being equal}, most households would choose to own more autos rather than less. It is therefore difficult to interpret these particular figures as anything but a confirmation of the expected.

The pure alternative preferences for the final group of ranked alternatives, housing type, are still more complex. No single alternative has zero utility, since all have either a constant term or some alternative specific socioeconomic variable associated with them. In general, the pure alternative utility of a housing type is a function of the household size and is conditional on auto ownership. The choice probabilities for the housing types are depicted in Figure 8.2. For each household size, four bar graphs representing the following household groups are presented:

(1) single worker households with less than two autos;
(2) single worker households with two or more autos;
(3) multi-worker households with less than two autos;
(4) multi-worker households with two or more autos.

In all these cases, only the constrained estimates are used.
Household Size

- Own single family
- Rent single family
- Rent garden-style or walk-up
- Rent high rise

*see text for explanation of codes.

Figure 8.2
PURE ALTERNATIVE HOUSING PROBABILITIES
As Figure 8.2 indicates, households which are small (i.e., two or fewer members) will generally select garden style or walk-up apartments with much higher probability than single family houses or high-rise apartments. However, this situation is completely reversed for larger households, where all else being equal, households seem to prefer single family units.

8.4 Value of Time

A traditional application of disaggregate models has been the development of estimates of value of time (e.g., Lisco, 1967). Typically, values of time have been derived by using the coefficient estimates of a utility function which includes measures of both cost and time. Thus, if IVTT, OVTT and OPTC are the in-vehicle time, out-of-vehicle time, and out-of-pocket cost respectively, and if the utility function for mode m is of the form*

\[ V_m = \Theta_1 IVTT_m + \Theta_2 OVTT_m + \Theta_3 OPTC_m + \ldots, \]

then the values of time implied are as follows:

\[
\text{Value of In-Vehicle Time} = -\frac{\delta V/\delta IVTT_m}{\delta V/\delta OPTC_m} = -\Theta_1/\Theta_3
\]

\[
\text{Value of Out-of-Vehicle Time} = -\frac{\delta V/\delta OVTT_m}{\delta V/\delta OPTC_m} = -\Theta_2/\Theta_3.
\]

Frequently, these models have included a term representing increasing value of time with increasing income (e.g., Lave, 1969). Many of these

*The notation \( \Theta \) is used to indicate model parameters in order to avoid confusion with particular \( \beta \)'s used in Chapter 7.
models also have other income related variables in the utility specification. The use of income in more than one variable gives rise to a serious theoretical question which is rarely addressed in the empirical literature. If income is a variable (denoted as before as $Y$) in the model, it is possible to derive different estimates of the value of time. For example, suppose the utility of the car mode is as follows:

$$V_c = \Theta_1 IVTT_c + \Theta_2 OVT_T + \Theta_4 Y + \ldots$$

then an alternative to the estimates of value of time given above (with appropriate adjustment for differences in units of measurement) is

$$\text{Value of In-Vehicle Time} = -\frac{\delta V}{\delta IVTT} / \frac{\delta V}{\delta Y} = -\frac{\Theta_1}{\Theta_4}$$

$$\text{Value of Out-of-Vehicle Time} = -\frac{\delta V}{\delta OVT} / \frac{\delta V}{\delta Y} = -\frac{\Theta_2}{\Theta_4}.$$ 

These estimates will quite naturally differ from those derived using out-of-pocket cost, and frequently differ by a great deal. For example, the value of in-vehicle time derived using the out-of-pocket cost variable in Peat, Marwick, and Mitchell's San Diego recommended modal split model from CBD oriented trips (PPM, 1972) is $3.18 per hour; however, if the income variable is used and an income of $10,000 per year is assumed, the value of time derived is nearly $432 per hour. Such differences can be attributed to the following sources:
(1) random variation - One of the key properties of many model estimation techniques (maximum likelihood included) is that the coefficient estimates they provide tend to be at least asymptotically distributed normally around the true coefficient. However, when a ratio of two such coefficient estimates is taken as in the value of time, the resulting estimate no longer has this and other desirable properties. The ratio of two normal variables has a highly skewed distribution with a generally high variance, which is a function of the variances of the numerator and denominator as well as their covariance and estimated means. Kendall and Stuart (1969) give the following formula as an approximation for the variance of the ratio of two normal estimates, \( \Theta_1 \) and \( \Theta_3 \).

\[
\text{Var}(\Theta_1/\Theta_3) = \left[ \frac{\text{E}(\Theta_1)/\text{E}(\Theta_3)}{\text{E}(\Theta_1)\text{E}(\Theta_3)} \right]^2 \left[ \text{Var}(\Theta_1)/\text{E}^2(\Theta_1) + \text{Var}(\Theta_3)/\text{E}^2(\Theta_3) - 2\text{Cov}(\Theta_1,\Theta_3)/\text{E}(\Theta_1)\text{E}(\Theta_3) \right].
\]

Ignoring any possible bias in the parameter estimates, the equation is simply

\[
\left[ \frac{\hat{\Theta}_1/\hat{\Theta}_3}{\hat{\Theta}_1/\hat{\Theta}_3} \right]^2 \left[ \text{Var}(\hat{\Theta}_1)/\hat{\Theta}_1 + \text{Var}(\hat{\Theta}_3)/\hat{\Theta}_3 - 2\text{Cov}(\hat{\Theta}_1,\hat{\Theta}_3)/\hat{\Theta}_1\hat{\Theta}_3 \right]
\]

where \( \hat{\Theta}_1 \) and \( \text{Var}(\hat{\Theta}_1) \) indicate estimates of \( \Theta_1 \) and its estimated variance respectively.
In this formula, even if the time and cost coefficients are estimated to have very low variance, their ratio typically has a very high variance, and is therefore not known with great precision.

(2) **differences in perception** - It is quite possible that decision-makers perceive income and direct out-of-pocket expenditures in quite a different way, and that a dollar spent on transit fares has greater disutility than a dollar decrease in income. In fact, differences in how monetary expenditures are perceived have long been recognized by businessmen who promote sales by encouraging the use of credit cards.

(3) **model mis-specification** - Specification error is always present to some degree in an econometric model; whether a model is useful simply depends on the degree of mis-specification. It is quite possible that income is measuring a variety of other factors not explicitly treated as independent variables. For example, income is correlated with life cycle, status, occupation and a host of other factors which might affect mobility choice. Mis-specifications such as these generally result in a bias in the coefficient estimate, and may be partly responsible for differences in value of time estimates.

In the joint mobility model developed in this study there is only one monetary term, the Z variable, which measures the income available to the household if it selected a given mobility bundle. Recalling the specification described in Section 7.3, the joint utility function can be written as follows:
$V = \ldots \beta_9 \text{TOTIME} + \beta_{10} \text{OVTT/DIST} + \beta_{11} \ln Z + \ldots$

The values of time (in dollars per hour) implicit in this functional form are therefore

Value of In-Vehicle Time $= -.24 Z \frac{\beta_9}{\beta_{11}}$.

Value of Out-of-Vehicle Time $= -.24 \left( \frac{\beta_9 + \beta_{10}}{\beta_{11} \text{DIST}} \right) Z$

where the factor of .24 is a correction for the units of Z and the time measures, and DIST is the two way work trip distance in miles. Note that the expression for the value of out-of-vehicle time reduces to the following:

Value of Out-of-Vehicle Time $(1 - .24 \frac{\beta_{10} Z}{\beta_{11} \text{DIST}})$.

Since $\beta_{10}$ is negative and $\beta_{11}$ is positive, the value of out-of-vehicle time in this functional form always exceeds the value of in-vehicle time, though the size of the difference varies inversely with the trip distance.

Note also that the value of out-of-vehicle time is a function of three parameter estimates, and its variance is still more complex than a simple ratio. To derive this variance, one must consider all the covariance terms as well as the trip distance.
Table 8.2 presents the unconstrained and constrained value of time estimates from the joint single worker household model for various values of Z and gives the estimated standard errors. A trip distance of 5 miles each way was assumed. As a rule of thumb, the value of Z is roughly one half of the household gross wage. Thus, a household earning $10,000 per year has an estimated value of in-vehicle time of $8 - $10 per hour, and a corresponding value for out-of-vehicle time of $10 - $16.

These estimates are somewhat higher than those frequently obtained, particularly for in-vehicle time. However, given their high standard error it is difficult to make any definitive statement about whether or not those differences are statistically significant. Furthermore, the previously discussed recent analysis of automobile ownership and mode to work by Lerman and Ben-Akiva produced value of time estimates which were in this range.

8.5 Prototypical Household Analysis: The Base Case

The limitation inherent in the previous analyses is that they focus on partial aspects of behavior such as the effect of a single variable or the household's time/money tradeoff without placing the predicted impacts in a realistic context. The approach adopted in this section is to construct a hypothetical city and examine the predicted mobility choice probabilities for a range of "prototypical" households. This method of analysis serves two purposes; it allows for a more intuitive presentation
<table>
<thead>
<tr>
<th>$Z$ ($/yr)</th>
<th>\textbf{UNCONSTRAINED ESTIMATES}</th>
<th></th>
<th>\textbf{CONSTRAINED ESTIMATES}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-Vehicle Time ($/hr)</td>
<td>S.E.</td>
<td>Out-of-Vehicle Time ($/hr)</td>
<td>S.E.</td>
</tr>
<tr>
<td>1000</td>
<td>1.86</td>
<td>1.15</td>
<td>3.14</td>
<td>2.42</td>
</tr>
<tr>
<td>3000</td>
<td>5.58</td>
<td>3.45</td>
<td>9.41</td>
<td>7.26</td>
</tr>
<tr>
<td>5000</td>
<td>9.30</td>
<td>5.75</td>
<td>15.68</td>
<td>12.10</td>
</tr>
<tr>
<td>10000</td>
<td>18.60</td>
<td>11.50</td>
<td>31.36</td>
<td>24.20</td>
</tr>
<tr>
<td>15000</td>
<td>27.90</td>
<td>17.25</td>
<td>47.04</td>
<td>36.30</td>
</tr>
</tbody>
</table>

Table 8.2

VALUES OF TIME
of the behavioral content of the models and provides a preliminary mechanism for predicting the potential impact of a variety of transportation policies.

This approach has the unfortunate disadvantage that many of the conclusions may be highly specific to the particular hypothetical city constructed. For this reason, an attempt was made to create a city which has attributes that are typical of older eastern cities. Much of the hypothetical data was actually selected from the Washington, D.C. base year data used in the model estimation. However, the reader is cautioned that many of the results may not be directly extendable to very different circumstances.

The hypothetical city is a simple two corridor, eleven zone city as depicted in Figure 8.3. The central business district, Zone 1, is assumed to be both an employment and residential site. The remaining 10 zones are structured in rings around the core. The rings are symmetric, and the distances from the zone centroids to the CBD are as follows:

<table>
<thead>
<tr>
<th>Zone</th>
<th>Distance to CBD (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2,3</td>
<td>4</td>
</tr>
<tr>
<td>4,5</td>
<td>7</td>
</tr>
<tr>
<td>6,7</td>
<td>12</td>
</tr>
<tr>
<td>8,9</td>
<td>15</td>
</tr>
<tr>
<td>10,11</td>
<td>25</td>
</tr>
</tbody>
</table>
The demographic characteristics of these zones are given in Table 8.3. For simplicity, the shopping generalized prices were assumed to be fixed for a zone rather than a function of the household income, and is therefore listed as a zonal attribute. The notation NA appears where the information is not applicable either because the alternative to which it applies is unavailable or, in the case of school expenditures, where the data was simply not needed.*

Each zone has attributes which are typical for different areas of existing U.S. cities. For example, Zone 2 might be termed a ghetto, i.e. a largely non-white residential area characterized by multi-family dwellings and low income. Zone 8 is an outer suburb with predominantly expensive owner-occupied single family homes, high average income, and is almost entirely white. The fringe zones, 10 and 11, consist entirely of low density owner-occupied dwellings.

State and local taxes were computed using 1968 Washington area rates. The zones were matched with Washington jurisdictions as follows:

<table>
<thead>
<tr>
<th>Zone</th>
<th>Jurisdiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3</td>
<td>District of Columbia</td>
</tr>
<tr>
<td>4,5</td>
<td>Alexandria, Va.</td>
</tr>
<tr>
<td>6,7</td>
<td>Fairfax City, Va.</td>
</tr>
<tr>
<td>8,9</td>
<td>Montgomery County, Md.</td>
</tr>
<tr>
<td>10,11</td>
<td>Loudoun County, Va.</td>
</tr>
</tbody>
</table>

* Zones 1, 2 and 3 are assumed to be in the District of Columbia, where school expenditures are not used in the model.
<table>
<thead>
<tr>
<th>Zone 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>630</td>
<td>630</td>
<td>750</td>
<td>650</td>
<td>890</td>
<td>725</td>
<td>700</td>
<td>700</td>
<td>Annual per pupil school expenditures</td>
</tr>
<tr>
<td>7500</td>
<td>5000</td>
<td>5000</td>
<td>9000</td>
<td>9000</td>
<td>11000</td>
<td>11000</td>
<td>15000</td>
<td>12000</td>
<td>15000</td>
<td>1400</td>
<td>Average annual income</td>
</tr>
<tr>
<td>.75</td>
<td>.98</td>
<td>.10</td>
<td>.50</td>
<td>.20</td>
<td>.20</td>
<td>.01</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
<td>.05</td>
<td>Fraction non-white</td>
</tr>
<tr>
<td>1.5</td>
<td>1.7</td>
<td>1.7</td>
<td>1.9</td>
<td>1.9</td>
<td>2.2</td>
<td>NA</td>
<td>2.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Transit shopping generalized price</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>50</td>
<td>25</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>.5</td>
<td>.5</td>
<td>Net residential density</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>100</td>
<td>100</td>
<td>400</td>
<td>400</td>
<td>900</td>
<td>900</td>
<td>500</td>
<td>500</td>
<td>Number of owner-occupied single family dwellings</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>100</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
<td>Number of rented single family dwellings</td>
</tr>
<tr>
<td>500</td>
<td>800</td>
<td>800</td>
<td>500</td>
<td>500</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>NA</td>
<td>NA</td>
<td>Number of garden-style or walk-up apartments</td>
</tr>
<tr>
<td>800</td>
<td>500</td>
<td>500</td>
<td>300</td>
<td>300</td>
<td>50</td>
<td>50</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Number of high rise apartments</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2500</td>
<td>2500</td>
<td>3500</td>
<td>30000</td>
<td>40000</td>
<td>30000</td>
<td>30000</td>
<td>30000</td>
<td>Average value of owner-occupied single family units</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>100</td>
<td>100</td>
<td>150</td>
<td>150</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td>NA</td>
<td>Monthly gross rent of rented single family dwellings</td>
</tr>
<tr>
<td>90</td>
<td>100</td>
<td>100</td>
<td>125</td>
<td>125</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td>NA</td>
<td>NA</td>
<td>Monthly gross rent of garden style &amp; high-rise apartments</td>
</tr>
<tr>
<td>90</td>
<td>80</td>
<td>80</td>
<td>100</td>
<td>100</td>
<td>120</td>
<td>120</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NS</td>
<td>Monthly gross rent of high-rise apartments</td>
</tr>
</tbody>
</table>

Table 8.3
ATTRIBUTES OF HYPOTHETICAL CITY
The transportation services in the hypothetical city are provided by conventional highway and bus systems. The highway network is identical in the two corridors, while the transit system is slightly better in the outer areas of even zones than in their odd counterparts. Table 8.4 presents the base case level of service measures to the downtown by both car and transit. Note that N.A. appears for transit measures in Zones 9 through 11 where transit is unavailable.

The model selected for this analysis is the constrained joint mobility choice model for single worker households. Using the unconstrained form violates the conditions for consistency in the grouping of alternatives established in Chapter 5, and would have resulted in forecasts which were highly sensitive to the particular zoning system selected.

In the base case, three household parameters were varied: income, race, and driver's license status. All the prototypical households used consist of a CBD worker, a spouse and two school-age children, neither of whom are old enough to have a license. Table 8.5 lists the twelve household types for which the choice probabilities in the base case were forecast. Households without licenses are basically captives to the transit system; zones without transit service are unavailable to them. All the other rules used to define the set of available alternatives in the model estimation were also invoked in the forecasting process.

Five basic impacts will be used to summarize the complete matrix of forecasted choice probabilities. These are:
<table>
<thead>
<tr>
<th>From</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car In-Vehicle Time (min.)</td>
<td>7.00</td>
<td>9.00</td>
<td>9.00</td>
<td>15.00</td>
<td>15.00</td>
<td>20.00</td>
<td>20.00</td>
<td>45.00</td>
<td>45.00</td>
<td>60.00</td>
<td>60.00</td>
</tr>
<tr>
<td>Car Out-of-Vehicle Time (min.)</td>
<td>9.00</td>
<td>9.00</td>
<td>9.00</td>
<td>8.00</td>
<td>8.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Car Out-of-Pocket Cost (c)</td>
<td>85.00</td>
<td>100.00</td>
<td>100.00</td>
<td>110.00</td>
<td>110.00</td>
<td>125.00</td>
<td>125.00</td>
<td>150.00</td>
<td>150.00</td>
<td>175.00</td>
<td>175.00</td>
</tr>
<tr>
<td>Transit In-Vehicle Time (min.)</td>
<td>7.00</td>
<td>15.00</td>
<td>15.00</td>
<td>20.00</td>
<td>25.00</td>
<td>30.00</td>
<td>40.00</td>
<td>60.00</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Transit Out-of-Vehicle Time (min.)</td>
<td>7.00</td>
<td>12.00</td>
<td>12.00</td>
<td>17.00</td>
<td>17.00</td>
<td>25.00</td>
<td>30.00</td>
<td>35.00</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Transit Fare (c)</td>
<td>25.00</td>
<td>25.00</td>
<td>25.00</td>
<td>25.00</td>
<td>50.00</td>
<td>50.00</td>
<td>75.00</td>
<td>90.00</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Figure 8.4
LEVEL OF SERVICE IN THE HYPOTHETICAL CITY
<table>
<thead>
<tr>
<th>Household</th>
<th>Income</th>
<th>Race</th>
<th>License Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LOW</td>
<td>WHITE</td>
<td>WITH</td>
</tr>
<tr>
<td>2</td>
<td>MIDDLE</td>
<td>WHITE</td>
<td>WITH</td>
</tr>
<tr>
<td>3</td>
<td>HIGH</td>
<td>WHITE</td>
<td>WITH</td>
</tr>
<tr>
<td>4</td>
<td>LOW</td>
<td>BLACK</td>
<td>WITH</td>
</tr>
<tr>
<td>5</td>
<td>MIDDLE</td>
<td>BLACK</td>
<td>WITH</td>
</tr>
<tr>
<td>6</td>
<td>HIGH</td>
<td>BLACK</td>
<td>WITH</td>
</tr>
<tr>
<td>7</td>
<td>LOW</td>
<td>WHITE</td>
<td>WITHOUT</td>
</tr>
<tr>
<td>8</td>
<td>MIDDLE</td>
<td>WHITE</td>
<td>WITHOUT</td>
</tr>
<tr>
<td>9</td>
<td>HIGH</td>
<td>WHITE</td>
<td>WITHOUT</td>
</tr>
<tr>
<td>10</td>
<td>LOW</td>
<td>BLACK</td>
<td>WITHOUT</td>
</tr>
<tr>
<td>11</td>
<td>MIDDLE</td>
<td>BLACK</td>
<td>WITHOUT</td>
</tr>
<tr>
<td>12</td>
<td>HIGH</td>
<td>BLACK</td>
<td>WITHOUT</td>
</tr>
</tbody>
</table>

Low = $3500/yr  
Middle = 9000/yr  
High = 22,500/yr  

Income Ranges

Table 8.5  
BASE CASE PROTotypical HOUSEHOLDS
(1) average distance to the CBD - one way in miles;

(2) housing type distribution-marginal choice probabilities for all four housing types;

(3) expected auto ownership;

(4) transit mode choice probability;

(5) total expected daily auto VMT (vehicle miles travelled).

The first measure summarizes the compactness of the expected residential pattern and is a useful mechanism for collapsing the entire location distribution to a single index. In addition, it reflects the expected work trip passenger miles travelled, a measure of the total transportation services which would have to be provided by the area.

The housing type choice probabilities represent the marginal probability distribution which summarizes the housing choices the household would make. The third measure, expected auto ownership, is a significant household characteristic upon which many non-work travel decisions are based, and is therefore of particular interest to the transportation planner. Similarly, the transit work mode choice probability is also a basic forecast needed for effective planning. The final measure, total expected work VMT by auto, is a statistic frequently used in determining the efficacy of various automotive pollution control strategies.

In addition to these four measures, any special effects which are relevant such as the choice probabilities for the ghetto area or the relative viability of the CBD as a residential area will be noted.
Table 8.6 is an impact summary for the twelve prototypical households in the base case. As might be anticipated, low income households are completely "priced out" of the owner-occupied single family housing market. Also, by definition households without licenses choose to own zero autos and take transit to work, so their expected auto ownership, transit mode choice probability and VMT are 0, 1 and 0 respectively.

Average work trip distance increases monotonically with income for all four race and drivers license status groups. Similarly, for households with licenses, increased income results in higher auto VMT. This result corresponds fairly well to the situation in real urban areas, in which the inner rings are for the most part lower income areas, where trips are short and auto usage is low.

The racial composition of the hypothetical city is such that non-white areas tend to lie fairly close to the CBD. For this reason, the work trip distances and VMT for black households are substantially smaller than for equivalent white households. This effect is eliminated in the high income households with drivers licenses, where expected VMT is actually somewhat higher for blacks than for whites.

Black households also tend to have a far lower probability of choosing single family ownership than do whites. This is partially because they tend to reside with relatively high probability in the ghetto area where there are no owner-occupied dwellings, and because two of the suburban
<table>
<thead>
<tr>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work Trip Distance</th>
<th>Own Single Family</th>
<th>Rent Single Family</th>
<th>Rent Gdn. or Walk-Up</th>
<th>Rent High Rise</th>
<th>Expected Auto Ownership</th>
<th>Transit Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>W</td>
<td>w/</td>
<td>6.42</td>
<td>0</td>
<td>.094</td>
<td>.764</td>
<td>.143</td>
<td>1.02</td>
<td>.409</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/</td>
<td>12.10</td>
<td>.299</td>
<td>.398</td>
<td>.280</td>
<td>.024</td>
<td>1.50</td>
<td>.181</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/</td>
<td>16.83</td>
<td>.676</td>
<td>.240</td>
<td>.080</td>
<td>.005</td>
<td>1.69</td>
<td>.084</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/</td>
<td>4.56</td>
<td>0</td>
<td>.061</td>
<td>.780</td>
<td>.159</td>
<td>.972</td>
<td>.451</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/</td>
<td>7.16</td>
<td>.138</td>
<td>.151</td>
<td>.642</td>
<td>.070</td>
<td>1.17</td>
<td>.296</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/</td>
<td>16.77</td>
<td>.692</td>
<td>.130</td>
<td>.163</td>
<td>.015</td>
<td>1.59</td>
<td>.086</td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/o</td>
<td>6.29</td>
<td>0</td>
<td>.258</td>
<td>.663</td>
<td>.079</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/o</td>
<td>7.64</td>
<td>.141</td>
<td>.309</td>
<td>.504</td>
<td>.045</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/o</td>
<td>11.50</td>
<td>.480</td>
<td>.292</td>
<td>.213</td>
<td>.015</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/o</td>
<td>5.00</td>
<td>0</td>
<td>.152</td>
<td>.749</td>
<td>.099</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/o</td>
<td>5.58</td>
<td>.053</td>
<td>.204</td>
<td>.674</td>
<td>.069</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/o</td>
<td>7.35</td>
<td>.157</td>
<td>.331</td>
<td>.469</td>
<td>.044</td>
<td>0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Income
L = Low
M = Middle
H = High

Race
W = White
B = Non-white

License
w/ = adults have license
w/o = adults lack license

Table 8.6
IMPACTS FOR BASE CASE
zones, 7 and 8, are 99% white.* This effect is somewhat reversed for high income blacks with drivers licenses, where the denial of opportunities in some relatively close suburbs (where there is some rental housing) forces them with disproportionate probability to the somewhat more integrated, exclusively owner-occupied fringe housing.

Black households also have lower auto ownership than their white counterparts, primarily because they are attracted to the ghetto which has good transit service. For similar reasons, blacks tend to select transit to work with higher probability than do whites. This effect also diminishes in the highest income group, where the relatively high probability of a black household choosing to live in the fringe area (where transit is unavailable) almost entirely balances the desirability of transit for those living near the CBD.

The group of households without licenses have a much more restricted set of alternatives. By definition, they must use transit to work and cannot own an automobile. Thus, zones 9, 10, and 11 are completely unavailable. Black households in this group must therefore live in zones 1 through 6. For this reason, expected work trip distance within the driverless households is extremely low.

The lack of a license also results in reducing the probability of owning a single family dwelling, since a large portion of such housing is

*By definition blacks are assigned a zero probability of choosing such zones.
outside the transit service boundary. A side effect of this is that the spatial distribution of those driverless households who would own their own home would be very limited. For example, white high income driverless households that did choose to own their home would select Zone 8 with probability .70 and equivalent black households would choose Zone 6 with probability .63.

The combination of being black, poor and lacking a license even further restricts available choices. Such households live in a zone which is fifty percent or more black (zones 1, 2 and 4) with probability .71 and have the shortest expected work trip distance. Interestingly, they tend to rent housing in multi-family dwelling units slightly less frequently than low income white households with licenses, since they have extra income such households might expend on auto ownership available for housing.

Within the city boundary (zone 1, 2 and 3), the expected residential pattern of black and white low income households is in sharp contrast. The location choice probabilities of zones 1 and 2 combined conditional on the city jurisdiction being selected is approximately .86 for low income black households with and without licenses, while it is only about .18 for low income whites.

8.6 Prototypical Household Analysis: Policy Scenarios

The base case discussed in the preceding section provides a useful reference point for evaluating the possible impacts of alternative trans-
portation policies. Six such policies, all designed to reduce automobile usage, will be considered. Automobile use reduction strategies were selected for evaluation because they are the major focus of current energy conservation and environmental protection programs.

*It should be made clear that the policy evaluation in this section is extremely limited in scope.* First, as in the analysis of the base case, it considers only the choice probabilities of individual households, not the aggregate group response. Further, it is only a demand analysis; the supply of housing, its prices, and other possible responses to demand changes are ignored. Finally, it is basically a long run analysis, in which the prototypical households may adjust their entire mobility bundle rather than simply one or two aspects of it.

For these reasons, the impacts forecasted should be viewed as "demand pressures" rather than as the resolution of a stable market equilibrium. For example, if due to a given policy the expected work trip length decreases, the real world short run impact may be an increase in housing prices near the CBD and a corresponding reduction near the fringe. In the longer run, new housing construction might be expected close to the center city. These changes will produce shifts in the demographic characteristics in the zone, which in turn will alter household location preferences. In short, the final aggregate response to a policy is determined by a
dynamic process which is far more complex than the demand forecasts presented here indicate, and all conclusions drawn from the analysis should be viewed as preliminary in nature. The actual numbers predicted are not intended to be actual forecasts and should not be used as such. Rather, they are summaries of disaggregate preferences for a given and fixed activity system.

The policies considered are designed to reflect various means of reducing car usage by either decreasing the attractiveness of using and owning automobiles or by improving the quality of transit service, thereby diverting auto users. The basic policies are as follows:

**Policy 1: Moderate auto use disincentives**

Impose a $1.00 parking charge downtown, increase auto operating costs by 50% via a fuel tax, increase auto out-of-vehicle time by 5 minutes per trip by regulating on-street parking.

**Policy 2: Strong auto use disincentives**

Take all actions in Policy 1 as well as a 25% auto ownership tax ($200 per year) and increase car in-vehicle time 25% by banning traffic in certain downtown areas.

**Policy 3: Moderate transit improvements**

Improve transit routing and scheduling to decrease in-vehicle time by 20% and out-of-vehicle time by 50%; halve all fares.

**Policy 4: Major transit improvements**

Install a new rapid transit system serving the entire urban area (including fringe) at speeds equal to that of the automobile in
non-CBD zones and 2 minutes faster in the CBD. System should have wait times 50% less than base case with a 15 minute maximum, and should have fares which are 50% of the base with a 25¢ maximum. Expected non-work generalized prices for shopping trips should be comparably improved.

**Policy 5: Moderate joint incentives**

Policies 1 and 3 combined.

**Policy 6: Major joint incentives**

Policies 2 and 4 combined.

The impact tables for these policies are presented in Tables 8.7 through 8.12. These tables are identical in structure to Table 8.6 which described the base case.

The first two policies which involve only changes to the attributes of car mode have no effect whatsoever on the choice probabilities of households without drivers licenses, since by definition they are captive to the transit mode. However, these policies do have a substantial impact on the decisions of the prototypical households with licenses. As expected, the transit marginal choice probabilities rise significantly from those in the base case. In absolute terms, this shift is greatest for the low income groups, ranging in Policy 1 from .188 to .084 for low and high income whites respectively. However, the greatest percentage shift, 89.4%, is for middle income blacks. Policy 2 has a similar though obviously greater effect on expected transit usage.
<table>
<thead>
<tr>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work trip Distance (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probability</th>
<th>Expected Total Daily VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>W</td>
<td>w/</td>
<td>6.01</td>
<td></td>
<td>.965</td>
<td>.597</td>
<td>5.47</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/</td>
<td>11.54</td>
<td>.245</td>
<td>.421</td>
<td>.308</td>
<td>.026</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/</td>
<td>16.77</td>
<td>.672</td>
<td>.243</td>
<td>.081</td>
<td>.005</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/</td>
<td>4.42</td>
<td>0</td>
<td>.051</td>
<td>.774</td>
<td>.176</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/</td>
<td>6.63</td>
<td>.103</td>
<td>.155</td>
<td>.670</td>
<td>.073</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/</td>
<td>16.69</td>
<td>.687</td>
<td>.132</td>
<td>.166</td>
<td>.015</td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/o</td>
<td>6.29</td>
<td>0</td>
<td>.258</td>
<td>.663</td>
<td>.079</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/o</td>
<td>7.64</td>
<td>.141</td>
<td>.309</td>
<td>.504</td>
<td>.045</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/o</td>
<td>11.50</td>
<td>.480</td>
<td>.292</td>
<td>.213</td>
<td>.015</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/o</td>
<td>5.00</td>
<td>0</td>
<td>.152</td>
<td>.749</td>
<td>.099</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/o</td>
<td>5.58</td>
<td>.053</td>
<td>.204</td>
<td>.674</td>
<td>.069</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/o</td>
<td>7.35</td>
<td>.157</td>
<td>.331</td>
<td>.469</td>
<td>044</td>
</tr>
</tbody>
</table>

**Income**
- L = Low
- M = Middle
- H = High

**Race**
- W = White
- B = Black

**License**
- w/ = adults have license
- w/o = adults lack license

Table 8.7
IMPACTS FOR POLICY 1
<table>
<thead>
<tr>
<th>Household</th>
<th>Worktrip Distance (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probability</th>
<th>Expected Total Daily VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Own Single Family</td>
<td>Rent Single Family</td>
<td>Rent Garden Apt.</td>
<td>Rent High Rise</td>
</tr>
<tr>
<td>L W w/</td>
<td>5.62</td>
<td>0</td>
<td>.081</td>
<td>.738</td>
<td>.006</td>
</tr>
<tr>
<td>M W w/</td>
<td>10.76</td>
<td>.179</td>
<td>.443</td>
<td>.348</td>
<td>.031</td>
</tr>
<tr>
<td>H W w/</td>
<td>16.31</td>
<td>.647</td>
<td>.260</td>
<td>.088</td>
<td>.005</td>
</tr>
<tr>
<td>L B w/</td>
<td>4.31</td>
<td>0</td>
<td>.048</td>
<td>.755</td>
<td>.197</td>
</tr>
<tr>
<td>M B w/</td>
<td>5.96</td>
<td>.065</td>
<td>.134</td>
<td>.704</td>
<td>.078</td>
</tr>
<tr>
<td>H B w/</td>
<td>16.01</td>
<td>.651</td>
<td>.142</td>
<td>.190</td>
<td>.018</td>
</tr>
<tr>
<td>L W w/o</td>
<td>6.29</td>
<td>0</td>
<td>.258</td>
<td>.663</td>
<td>.079</td>
</tr>
<tr>
<td>M W w/o</td>
<td>7.64</td>
<td>.141</td>
<td>.309</td>
<td>.504</td>
<td>.045</td>
</tr>
<tr>
<td>H W w/o</td>
<td>11.50</td>
<td>.480</td>
<td>.292</td>
<td>.213</td>
<td>.015</td>
</tr>
<tr>
<td>L B w/o</td>
<td>5.00</td>
<td>0</td>
<td>.152</td>
<td>.747</td>
<td>.099</td>
</tr>
<tr>
<td>M B w/o</td>
<td>5.58</td>
<td>.053</td>
<td>.204</td>
<td>.674</td>
<td>.069</td>
</tr>
<tr>
<td>H B w/o</td>
<td>7.35</td>
<td>.157</td>
<td>.331</td>
<td>.469</td>
<td>.044</td>
</tr>
</tbody>
</table>

**Income**
- L = Low
- M = Middle
- H = High

**Race**
- W = White
- B = Black

**License**
- w/ = adults have license
- w/o = adults lack license

Table 8.8
IMPACTS FOR POLICY 2
<table>
<thead>
<tr>
<th>Household</th>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work trip Distance (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probabilities</th>
<th>Expected Total VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Own Single Family</td>
<td>Rent Single Family</td>
<td>Rent Garden Apt.</td>
<td>Rent High Rise</td>
</tr>
<tr>
<td>L W w/</td>
<td>7.06</td>
<td></td>
<td>0</td>
<td>.119</td>
<td>.744</td>
<td>.136</td>
<td>1.00</td>
<td>5.01</td>
</tr>
<tr>
<td>M W w/</td>
<td>11.94</td>
<td></td>
<td>.291</td>
<td>.403</td>
<td>.282</td>
<td>.024</td>
<td>1.49</td>
<td>2.246</td>
</tr>
<tr>
<td>H W w/</td>
<td>16.62</td>
<td></td>
<td>.664</td>
<td>.248</td>
<td>.083</td>
<td>.005</td>
<td>1.69</td>
<td>1.122</td>
</tr>
<tr>
<td>L B w/</td>
<td>4.58</td>
<td></td>
<td>0</td>
<td>.068</td>
<td>.778</td>
<td>.154</td>
<td>.965</td>
<td>.524</td>
</tr>
<tr>
<td>M B w/</td>
<td>7.01</td>
<td></td>
<td>.130</td>
<td>.152</td>
<td>.648</td>
<td>.070</td>
<td>1.15</td>
<td>.351</td>
</tr>
<tr>
<td>H B w/</td>
<td>16.53</td>
<td></td>
<td>.678</td>
<td>.135</td>
<td>.172</td>
<td>.016</td>
<td>1.57</td>
<td>.110</td>
</tr>
<tr>
<td>L W w/o</td>
<td>6.52</td>
<td></td>
<td>0</td>
<td>.280</td>
<td>.645</td>
<td>.076</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M W w/o</td>
<td>7.91</td>
<td></td>
<td>.157</td>
<td>.317</td>
<td>.483</td>
<td>.043</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>H W w/o</td>
<td>11.70</td>
<td></td>
<td>.496</td>
<td>.289</td>
<td>.201</td>
<td>.014</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>L B w/o</td>
<td>5.11</td>
<td></td>
<td>0</td>
<td>.163</td>
<td>.740</td>
<td>.097</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M B w/o</td>
<td>5.70</td>
<td></td>
<td>.058</td>
<td>.213</td>
<td>.663</td>
<td>.067</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>H B w/o</td>
<td>7.50</td>
<td></td>
<td>.165</td>
<td>.338</td>
<td>.455</td>
<td>.042</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Income**
L = Low  
M = Middle  
H = High

**Race**
W = White  
B = Black

**License**
w/ = adults have licence  
w/o = adults lack license

Table 8.9  
IMPACTS FOR POLICY 3
<table>
<thead>
<tr>
<th>Household</th>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work Distance (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probabilities</th>
<th>Expected Total Daily VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Own Single Family</td>
<td>Rent Single Family</td>
<td>Rent Garden Apt.</td>
<td>Rent High Rise</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/</td>
<td>5.72</td>
<td>0</td>
<td>.137</td>
<td>.733</td>
<td>.130</td>
<td>.925</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/</td>
<td>7.90</td>
<td>.128</td>
<td>.307</td>
<td>.517</td>
<td>.049</td>
<td>1.01</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/</td>
<td>11.90</td>
<td>.442</td>
<td>.333</td>
<td>.210</td>
<td>.016</td>
<td>.972</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/</td>
<td>4.43</td>
<td>0</td>
<td>.070</td>
<td>.775</td>
<td>.156</td>
<td>.951</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/</td>
<td>5.13</td>
<td>.035</td>
<td>.147</td>
<td>.737</td>
<td>.082</td>
<td>1.05</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/</td>
<td>9.10</td>
<td>.181</td>
<td>.276</td>
<td>.494</td>
<td>.050</td>
<td>1.03</td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/o</td>
<td>7.39</td>
<td>0</td>
<td>.328</td>
<td>.604</td>
<td>.068</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/o</td>
<td>10.75</td>
<td>.327</td>
<td>.299</td>
<td>.345</td>
<td>.030</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/o</td>
<td>15.13</td>
<td>.664</td>
<td>.214</td>
<td>.115</td>
<td>.007</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/o</td>
<td>5.21</td>
<td>0</td>
<td>.166</td>
<td>.738</td>
<td>.095</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/o</td>
<td>6.75</td>
<td>.120</td>
<td>.204</td>
<td>.615</td>
<td>.062</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/o</td>
<td>11.88</td>
<td>.419</td>
<td>.245</td>
<td>.309</td>
<td>.028</td>
<td></td>
</tr>
</tbody>
</table>

**Income**
- L = Low
- M = Middle
- H = High

**Race**
- W = White
- B = Black

**License**
- w/ = adults have license
- w/o = adults lack license

Table 8.10
IMPACTS FOR POLICY 4
<table>
<thead>
<tr>
<th>Household</th>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work trip (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probabilities</th>
<th>Expected Total Daily VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Own Single Family</td>
<td>Rent Single Family</td>
<td>Rent Garden Apt.</td>
<td>Rent High Rise</td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/</td>
<td>6.21</td>
<td>0</td>
<td>.120</td>
<td>.731</td>
<td>.149</td>
<td>.954</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/</td>
<td>11.40</td>
<td>.240</td>
<td>.426</td>
<td>.307</td>
<td>.027</td>
<td>1.45</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/</td>
<td>16.53</td>
<td>.658</td>
<td>.253</td>
<td>.084</td>
<td>.005</td>
<td>1.68</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/</td>
<td>4.48</td>
<td>0</td>
<td>.062</td>
<td>.772</td>
<td>.166</td>
<td>.941</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/</td>
<td>6.50</td>
<td>.096</td>
<td>.156</td>
<td>.674</td>
<td>.073</td>
<td>1.12</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/</td>
<td>16.41</td>
<td>.670</td>
<td>.138</td>
<td>.176</td>
<td>.016</td>
<td>1.56</td>
</tr>
<tr>
<td>L</td>
<td>W</td>
<td>w/o</td>
<td>6.52</td>
<td>0</td>
<td>.280</td>
<td>.645</td>
<td>.076</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>W</td>
<td>w/o</td>
<td>7.91</td>
<td>.157</td>
<td>.317</td>
<td>.483</td>
<td>.043</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>w/o</td>
<td>11.70</td>
<td>.496</td>
<td>.289</td>
<td>.201</td>
<td>.014</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>B</td>
<td>w/o</td>
<td>5.11</td>
<td>0</td>
<td>.163</td>
<td>.740</td>
<td>.097</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td>w/o</td>
<td>5.70</td>
<td>.058</td>
<td>.213</td>
<td>.663</td>
<td>.067</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>B</td>
<td>w/o</td>
<td>7.50</td>
<td>.165</td>
<td>.338</td>
<td>.455</td>
<td>.042</td>
<td>0</td>
</tr>
</tbody>
</table>

Income: 
- **L** = Low
- **M** = Middle
- **H** = High

Race: 
- **W** = White
- **B** = Black

License: 
- **w/** = adults have license
- **w/o** = adults lack license

Table 8.11

IMPACTS FOR POLICY 5
<table>
<thead>
<tr>
<th>Household</th>
<th>Income</th>
<th>Race</th>
<th>License</th>
<th>Work trip distance (miles)</th>
<th>Housing Choice Probabilities</th>
<th>Expected Auto Ownership</th>
<th>Transit Choice Probabilities</th>
<th>Expected Total Daily VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>L W w/</td>
<td>5.66</td>
<td>0</td>
<td>.119</td>
<td>.724</td>
<td>.157</td>
<td>.875</td>
<td>.788</td>
<td>2.15</td>
</tr>
<tr>
<td>M W w/</td>
<td>7.78</td>
<td>.114</td>
<td>.307</td>
<td>.528</td>
<td>.051</td>
<td>.991</td>
<td>.504</td>
<td>7.50</td>
</tr>
<tr>
<td>H W w/</td>
<td>11.84</td>
<td>.435</td>
<td>.336</td>
<td>.214</td>
<td>.016</td>
<td>.959</td>
<td>.463</td>
<td>12.44</td>
</tr>
<tr>
<td>L B w/</td>
<td>4.37</td>
<td>0</td>
<td>.060</td>
<td>.757</td>
<td>.183</td>
<td>.899</td>
<td>.763</td>
<td>1.92</td>
</tr>
<tr>
<td>M B w/</td>
<td>5.10</td>
<td>.032</td>
<td>.146</td>
<td>.739</td>
<td>.083</td>
<td>1.04</td>
<td>.498</td>
<td>5.09</td>
</tr>
<tr>
<td>H B w/</td>
<td>7.05</td>
<td>.178</td>
<td>.277</td>
<td>.495</td>
<td>.050</td>
<td>1.02</td>
<td>.463</td>
<td>8.16</td>
</tr>
<tr>
<td>L W w/o</td>
<td>7.39</td>
<td>0</td>
<td>.328</td>
<td>.604</td>
<td>.068</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M W w/o</td>
<td>10.75</td>
<td>.327</td>
<td>.299</td>
<td>.345</td>
<td>.030</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>H W w/o</td>
<td>15.13</td>
<td>.664</td>
<td>.214</td>
<td>.115</td>
<td>.007</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>L B w/o</td>
<td>5.21</td>
<td>0</td>
<td>.166</td>
<td>.738</td>
<td>.095</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M B w/o</td>
<td>6.75</td>
<td>.120</td>
<td>.204</td>
<td>.615</td>
<td>.062</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>H B w/o</td>
<td>11.88</td>
<td>.419</td>
<td>.245</td>
<td>.309</td>
<td>.028</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Income**  
L = Low  
M = Middle  
H = High  

**Race**  
W = White  
B = Black  

**License**  
w/ = adults have license  
w/o = adults lack license  

**Table 8.12**  
IMPACTS FOR POLICY 6
Policies 1 and 2 have a much more limited impact on automobile ownership. For the high income households the greatest effect is about .5% and 3% for whites and blacks respectively, and is never greater than 13% for any household.

The first two policies have only a marginal impact on the expected spatial distribution of households as measured by the average work trip distance. However, the large changes in transit usage coupled with the small reductions in average trip length result in a substantial reduction in expected auto VMT. This is particularly noticeable in the lowest income households, for whom expected VMT more than halves in Policy 2.

In the first two policies, the biggest shifts in the housing decision are for middle income households. Increased car costs tend to make home ownership less attractive, both because less income is available for such purposes and because suburban housing is now associated with higher car commuting costs. In a probabilistic sense, it appears that when faced with higher auto costs, middle income households shift their housing type far more readily than they alter their level of auto ownership. This effect is minimal for higher income households, and doesn't exist for low income households, for whom home ownership is not an available alternative.

Policy 3, which involves moderate transit improvements, in general produces much smaller changes than either Policy 1 or 2. However, it does affect both households with and without licenses. The overall impact can be characterized by a reasonably large shift to the transit mode and only
very marginal changes in all other impacts. Average trip length for low income households and those without licenses is slightly increased, while it is decreased for middle and upper income households with licenses. Thus, the moderately improved transit system allows the low income households and those captive to transit to select further zones with higher probability and provides some moderate incentives for more affluent households to move inward.

The fourth policy represents radical changes in the public transportation system. As one might expect, such changes have very significant impacts on virtually all aspects of the mobility decision. The extension of the public transportation system to cover the entire metropolitan area opens new locational opportunities for households without licenses, and the high level of service the system offers coupled with low fares permits low income groups to move further from the CBD. All these factors are reflected in Table 8.10.

Some of the effects forecast for Policy 4 are not intuitively obvious. First, average work trip distances for households without licenses are actually greater than for corresponding households with licenses. This is primarily caused by a difference in the income allocation of these households; households without licenses by definition do not own cars and therefore allocate more income to the purchase of better housing, which tends to lie in the outer suburban areas of the hypothetical city. This factor is also reflected in the large reduction in the choice probabilities for
owner-occupied single family dwellings for households with licenses, and a corresponding increase for those lacking licenses.

Another interesting effect of Policy 4 is the reduction in average automobile ownership for high income groups. For both black and white households, the expected auto ownership for high income households is actually less than for the corresponding middle income households. This results from the interaction between the location and auto ownership decisions. Since high income households have strong preferences for not residing in lower income areas, they tend to locate in zones of high average income. In the base case and the previous policies, these zones had poor or non-existent transit and hence were associated with high auto ownership; however, under Policy 4 those zones are extremely well served by public transportation, and the high income households residing there allocate more of their income to housing and non-mobility related expenditures and less to the purchase of automobiles.

Expected VMT in Policy 4 is still a monotonic function of income. However, the expected VMT of low income households is actually reduced by more than a factor of two from the base case.

Policies 5 and 6, which represent combinations of auto use disincentives and transit improvements at moderate and high levels respectively, produce impacts which are similar to the previously described policies. The distribution of the impacts of the moderate policy (Policy 5) seems to
be weighted quite heavily towards low income households while the major transportation system changes (Policy 6) produce strong impacts across all socioeconomic groups.

Of all the policies considered, only Policies 4 and 6 produce profound changes in the current location, housing, auto ownership and mode to work patterns. The remaining policies all result in patterns which are variants of the base case, i.e., residential patterns characterized by low income groups in the inner city with low auto ownership and more affluent residents in the suburbs with high auto ownership. This seems to indicate that while the public transportaion policy can have significant and widespread effects on locational preferences and is potentially an important policy instrument for altering urban form, major changes rather than incremental ones will be required.

8.7 Summary of Behavioral Implications

The models developed in this study can be used in a variety of ways to draw inferences about household mobility behavior; each evaluation approach is in some way limited, and valid conclusions must ultimately be synthesized from a composite of the approaches used.

There appear to be significant and explainable differences between the behavior of single and multi-worker households as reflected in the parameter estimates of their utility functions. This behavioral variation probably arises from important differences in the life styles of these two groups. Single worker households seem to place more weight on neighbor-
hood characteristics than do their multi-worker counterparts. Also, additional drivers in single worker households have a much greater marginal effect on auto ownership than in equivalent multi-worker households. This latter difference probably is the result of the potential for two or more workers within a household to share an auto for making work trips, while in single worker households use of an auto by one licensed member generally excludes its use by others.

All else being equal, housing preferences for both single and multi-worker households seem to be significantly affected by household size. Large households have strong preferences for single family dwellings, while small ones appear to prefer apartments.

The point estimates of the values of time implied by the single worker household models are somewhat higher than has been found in conventional mode choice models, though they are consistent with more recent analysis by Lerman and Ben-Akiva (1975). However, these estimates are subject to extremely high variance, and it is therefore difficult to draw inferences about their true values.

Forecasts of the choice probabilities for a number of single worker households in an hypothetical city provide some significant insights into the possible impact of alternative transportation policies. These policies do not have uniform impacts on all socioeconomic groups. Of the policies considered, those directed towards discouraging auto use by decreasing car level of service impacted most heavily on the poorest households, and did
not affect households without drivers licenses. However, policies which produced substantial improvements in the public transit system did have major effects on all segments of society by opening up new opportunities to transit "captive" and providing powerful incentives for a more consolidated location pattern for affluent households.
Chapter 9
CONCLUSIONS AND RECOMMENDATIONS

9.1 Summary of the Thesis

The study described in the thesis represents the first attempt to develop a joint disaggregate model of household location, housing, auto ownership and mode to work decisions which is rooted in a consistent theory of consumer behavior, uses household level data, and deals with the entire set of relevant choices in a joint decision framework. This approach has permitted an explicit representation of a broad range of causal factors and effectively captured the most significant tradeoffs the household can make in its long run transportation and location decisions.

The theoretical basis for this model is that the choice of location, housing, auto ownerships, and mode to work, termed the mobility decision, is made in a substantially longer time frame than day-to-day non-work travel choices of frequency, destination, mode, time of day, and route. Thus, the principal assumption of this thesis is that the entire spectrum of travel-related choices can be behaviorally partitioned into long run mobility decisions, which are the focus of this study, and short run travel choices. These decisions are structured hierarchically; travel decisions are made conditional on the outcome of the mobility choices.
A variety of hypotheses about the factors which might affect the mobility decision were considered. These factors include travel time, spatial opportunities, urban services, taxes, pollution, neighborhood quality, race and ethnicity as well as the socioeconomic characteristics of the household. In order to consider the effects of these factors on the mobility decision, the classical theory of household bid rent curves was extended to include each causal effect. This approach permitted a graphical analysis of alternative hypotheses about how households perceive the various factors which impinge on their mobility choice.

The basic methodology selected for the empirical study is the multinomial logit model. The models were estimated using cross-sectional household level data from the 1968 Washington, D.C. Home Interview survey augmented primarily by U.S. Census data. This data base, while somewhat lacking in terms of its description of the housing choice, is representative of the type of data likely to be available in many major U.S. cities. Hence, the study approach should be readily extendable to other urban areas.

Separate models were estimated for single and multiple worker households. This was done because the underlying mobility decision processes of these two groups were hypothesized on a priori grounds to be very different in nature. Overall, the empirical results for the single worker households are far superior to those for multi-worker families. This is partially due to the apparent complexity inherent in the way in which the
proximity of a location to the workplaces of more than one worker is evaluated by the household. Attempts at simplifying this process all proved unsuccessful and the final models selected represent a compromise between computational convenience and the need to consider the attributes of all work trips in a household jointly.

The estimated models were then evaluated to develop behavioral insights into the mobility choice process. Coefficients from the single and multi-worker household models were compared, the pure alternative effects which measure the relative desirability of a mobility alternative under rigid ceterus paribus conditions were considered, the values of time implied by the models were derived, and choice probabilities for a variety of prototypical households making mobility decisions in an hypothetical city were forecast. In the last of these evaluations, the potential impacts of a range of alternative transportation strategies designed to reduce the use of the private automobile were considered.

9.2 Implications of the Research for a Revised Analysis Framework

Forecasts of urban land use have traditionally played an important role in the transportation planning process. However, the models used to forecast residential and employment location patterns have usually been logically separable from those used to forecast both automobile ownership and trip-making patterns. This traditional process is depicted in Figure 9.1. Land use models provide forecasts of zonal population and employment. A separate auto ownership model is then applied. These
Figure 9.1
THE TRADITIONAL FORECASTING PROCESS
forecasts are then used as inputs to the "four-step process," consisting of trip generation, distribution, mode split and assignment, usually applied for three trip purposes (home based work, home based non-work and home based other). As indicated in the figure, feedback between the various steps is rarely incorporated into the analysis, though this is more a shortcoming of current practice rather than an intrinsic property of the model system.

A critical implication of the analysis in this study is that such a forecasting approach fails to behaviorally represent the true process it seeks to model. In reality, both automobile ownership and work trip travel patterns arise as a logical consequence of the mobility choice process. It therefore seems reasonable that auto ownership and work trip patterns should be forecast within what has traditionally been termed urban land use forecasting. For work trips, trip generation actually represents labor force participation, a decision process which probably depends more on the household structure, life cycle, and the state of the regional economy than on the transportation system.* Work trip distribution is simply the outcome of urban location patterns rather than a distinct behavioral phenomenon to be modelled separately.

The use of sequential model structures in travel forecasting is also behaviorally unsound. In behavioral terms, the decision process being represented is actually a joint one. Thus, unless a sequential model

*It may be possible that labor force participation is actually dependent on access to employment, though this has yet to be convincingly demonstrated.
structure is simply a computationally convenient means of mathematically decomposing a model estimated as a joint process into a series of forecasting steps (Manheim, 1973), a joint forecasting process should be adopted.

A revised forecasting approach is depicted in Figure 9.2. In this diagram, work trip forecasts arise as a direct consequence of the land use forecasts, while non-work travel is the result of a joint travel choice process for each trip purpose. This forecasting approach is a direct aggregate analogue of the choice hierarchy first proposed by Ben-Akiva (1973) and discussed in Chapter 2, except that a supply sector and appropriate feedback have been added to resolve both the long run equilibrium in land use and the short run equilibrium in the network assignment. *

The mobility choice models developed in this thesis are only one part of the system of models displayed in Figure 9.2. However, many of the other model components are the object of a great deal of research. Joint travel choice models are being developed by Adler and Ben-Akiva (1975) and Ben-Akiva and Richards (1975). New models of the land use supply sector are being studied by the National Bureau of Economic Research (1974), and new equilibration techniques are being developed by Cambridge Systematics (1974). Thus, the proposed forecasting process represents a synthesis of the outcome of ongoing research in a variety of areas, and may be implementable in a fairly short period of time.

*It is not essential that equilibrium actually exist, as long as the supply/demand interaction is modelled.
Figure 9.2

A REVISED FORECASTING PROCESS
A critical and still unresolved issue in applying disaggregate choice models in this framework is the aggregation problem, i.e. the way in which aggregate forecasts can be made from disaggregate models. Theoretically, the solution to this problem is straightforward; aggregate behavior is the sum of the choices of individuals. However, in practical terms, the summation of individual behavior would require knowledge of the complete distribution of independent variables for all zones for which forecasts were to be made.

When aggregate land use models are used, such information is generally lost for later use in travel demand forecasting. However, disaggregate behavioral mobility choice models such as those described in this thesis offer the potential for providing much of the needed distributional data. This can be done through the random sample aggregation method (Koppelman, 1975).

In the random sample aggregation method, a group of households are used for the entire forecasting process. The choice probabilities for each household are forecast at each step of the process, and aggregate forecasts are made by simply applying the appropriate sampling factors to the probabilities. Thus, the original sample of households and their choice probabilities represent most of the entire distribution of independent variables.*

*The problem of within zone variance of level of service characteristics still remains. However, work by Talvitie (1974) addresses precisely this problem.
This forecasting method has a number of advantages over aggregation approaches such as those of Talvitie (1973), Westin (1974), and McFadden (1975):

(1) It does not require any assumptions about the particular form of the distribution of the independent variables such as the assumed normality of Westin's and the McFadden/Reid approach. The random sample of households, if sufficiently large, can adequately represent any discrete or continuous distribution.

(2) No analytic simplifications such as the truncation of higher order moments of the distributions implicit in Talvitie's method are required.

(3) The computational requirements of the forecasting process can be directly controlled by varying the sample size. If more accurate forecasts are desired, larger samples can be used.

(4) The relative forecasting accuracy for particular socioeconomic groups can be regulated. If the choices of one group such as the poor, the elderly, or the handicapped are of particular interest then a relatively large sample of households with such persons can be used.

(5) Since disaggregate models are used throughout the forecasting process, the number of observations required can be greatly reduced. The use of aggregate models results in a substantial loss of information, which is avoided in disaggregate modelling methodologies. The current emphasis on the quantity of data collected can be shifted to upgrade the reliability of the data by improving checking and verification methods.

9.3 Recommendations for Future Research

As in most empirical research, this study probably raises more questions than it answers. A range of important research problems must ultimately be addressed if the analysis methodology proposed in the preceding section is to become operational. Some of the most significant of these research topics are as follows:
(1) The joint mobility choice models estimated in this study should be re-estimated with significantly larger samples to obtain more reliable coefficient estimates. This will permit more precise behavioral inferences to be drawn from the results.

(2) Data from other U.S. cities should be used for estimation in order to test the extent to which the results are geographically and temporally transferable and whether differences in the coefficients across cities or time periods can be explained in terms of intrinsic differences in the physical or demographic structure of the cities. Recent work by Atherton (1975) using mode choice models tends to indicate that a great deal of geographical transferability does exist, and that data requirements for model estimation can be substantially reduced by using prior estimates in a Bayesian parameter updating procedure. This work should be extended to models of more complex choice situations to test the extent to which transferability exists.

(3) Attempts should be made to improve the specification of the models, particularly those for multi-worker households. The most critical issue in the development of these models is the way in which the preferences of the workers within a household interact in the location decision. Do all workers have equal weights in the decision-making process, or are their relative priorities determined by their income, status in the household, or other factors? This analytic research should be supplemented by a more attitudinally oriented study to permit an in-depth examination of how multi-worker households reach their location decisions. This work should provide more concrete hypotheses about the types of model specifications which will prove useful in an empirical study.

(4) Further experiments with estimating models for different socio-economic groups, or market segments, should be undertaken. The study of auto ownership and mode to work by Lerman and Ben-Akiva (1975) indicates that both life cycle and occupation groups may prove to be useful for market segmentation. This research should be extended to the full mobility choice. Further exploration into homogeneity of preferences across racial and income groups should be undertaken.
The analysis of prototypical households described in Chapter 8 should be extended to make aggregate level forecasts, both in hypothetical and real cities. These forecasts will provide a useful tool for general policy analysis and yield further insight into the behavioral structure of mobility decisions.

In the longer term, attempts should be made to integrate improved versions of the mobility choice demand models developed in this study with a housing supply model. The current version of the NBER Urban Simulation Model is probably the most useful base for such an effort. However, a broad range of difficult equilibration problems will have to be explored. Currently, prices for each time period are determined as a function of the shadow prices from a linear programming market resolution process in previous time periods. This process might be replaced by some type of probabilistic simulation-oriented household assignment in which households are incrementally "loaded" onto the supply of housing. However, it is unclear how market prices will be obtained in this framework.

It may prove feasible to apply disaggregate choice theory to modelling the location decisions of centers of population-serving employment. Firms, like households, have specific objectives in choosing their location. These objectives relate principally to a perception of the potential profitability of alternative sites. Like households, firms have particular characteristics such as their size and products which affect their locational preferences. Given the success of choice theory in modelling consumer behavior, at least some attempt should be made to collect appropriate data and estimate an employer location model.

Work should begin on restructuring the urban transportation planning process to apply the random sample aggregation method with joint, behavioral disaggregate models. The effectiveness of this approach should be first verified in a case study in an actual city, and computer software should be designed and implemented to permit computationally efficient forecasts to be made. Ultimately, direct comparisons in terms of cost, accuracy, data requirements and flexibility between the current forecasting process, various modifications to it which retain the sequential structure but use disaggregate methodology and the proposed random sample aggregation method should be made.
9.4 Conclusion

The interaction between urban form and the transportation system may be the most critical factor in determining the efficacy of long term transportation strategies. However, in the past decade the principal advances in transportation modelling have been in the field of travel demand, in which the pattern of residential location has been assumed to be given. However, even the best travel demand models are virtually useless for policy analysis unless they are coupled with much more effective land use forecasts than can currently be provided.

This study represents one step in the process of improving land use forecasts by introducing joint behavioral choice models into the representation of the household mobility choice. Obviously, this study must be one portion of a much larger research effort to restructure the entire travel forecasting process to better reflect a behavioral understanding of the true causal mechanisms which determine both supply and demand. Only by developing more behaviorally structured models can transportation analysts and urban planners hope to provide reliable forecasts of the impact of alternative policies upon which an informed decision-making process can act.
BIBLIOGRAPHY


Laurenti, Luigi M. Property Values and Race, University of California Press, Berkeley, California, 1960.


von Thunen, J. *Der Isolierte Staat*. 1842


