REPORT S2-C04-RW-1

Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand

SHRP2 CAPACITY RESEARCH

SHRP 2
STRATEGIC HIGHWAY RESEARCH PROGRAM
Accelerating solutions for highway safety, renewal, reliability, and capacity

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

The Second

STRATEGIC HIGHWAY RESEARCH PROGRAM

SHRP 2 REPORT S2-C04-RW-1

Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand

Parsons Brinckerhoff

Northwestern University

Mark Bradley Research & Consulting

University of California at Irvine

Resource System Group

University of Texas at Austin

Frank Koppelman, and

GeoStats

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.

2013

www.TRB.org

Subscriber Categories

Highways
Planning and Forecasting

The Second Strategic Highway Research Program

America's highway system is critical to meeting the mobility and economic needs of local communities, regions, and the nation. Developments in research and technology—such as advanced materials, communications technology, new data collection technologies, and human factors science—offer a new opportunity to improve the safety and reliability of this important national resource. Breakthrough resolution of significant transportation problems, however, requires concentrated resources over a short time frame. Reflecting this need, the second Strategic Highway Research Program (SHRP 2) has an intense, large-scale focus, integrates multiple fields of research and technology, and is fundamentally different from the broad, mission-oriented, discipline-based research programs that have been the mainstay of the highway research industry for half a century.

The need for SHRP 2 was identified in *TRB Special Report 260: Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life,* published in 2001 and based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP 2, modeled after the first Strategic Highway Research Program, is a focused, time-constrained, management-driven program designed to complement existing highway research programs. SHRP 2 focuses on applied research in four areas: Safety, to prevent or reduce the severity of highway crashes by understanding driver behavior; Renewal, to address the aging infrastructure through rapid design and construction methods that cause minimal disruptions and produce lasting facilities; Reliability, to reduce congestion through incident reduction, management, response, and mitigation; and Capacity, to integrate mobility, economic, environmental, and community needs in the planning and designing of new transportation capacity.

SHRP 2 was authorized in August 2005 as part of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). The program is managed by the Transportation Research Board (TRB) on behalf of the National Research Council (NRC). SHRP 2 is conducted under a memorandum of understanding among the American Association of State Highway and Transportation Officials (AASHTO), the Federal Highway Administration (FHWA), and the National Academy of Sciences, parent organization of TRB and NRC. The program provides for competitive, merit-based selection of research contractors; independent research project oversight; and dissemination of research results.

SHRP 2 Report S2-C04-RW-1

ISBN: 978-0-309-12969-5

Copyright Information

Note: SHRP 2 report numbers convey the program, focus area, project number, and publication format. Report numbers ending in "w" are published as web documents only.

Notice

The project that is the subject of this report was a part of the second Strategic Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical committee and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

SHRP 2 Reports

Available by subscription and through the TRB online bookstore:
www.TRB.org/bookstore

Contact the TRB Business Office:
202-334-3213

More information about SHRP 2:
www.TRB.org/SHRP2

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. (Dan) Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. (Dan) Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

SHRP 2 STAFF

Ann M. Brach, *Director*
Stephen J. Andrle, *Deputy Director*
Neil J. Pedersen, *Deputy Director, Implementation and Communications*
James Bryant, *Senior Program Officer, Renewal*
Kenneth Campbell, *Chief Program Officer, Safety*
JoAnn Coleman, *Senior Program Assistant, Capacity and Reliability*
Eduardo Cusicanqui, *Financial Officer*
Walter Diewald, *Senior Program Officer, Safety*
Jerry DiMaggio, *Implementation Coordinator*
Shantia Douglas, *Senior Financial Assistant*
Charles Fay, *Senior Program Officer, Safety*
Carol Ford, *Senior Program Assistant, Renewal and Safety*
Elizabeth Forney, *Assistant Editor*
Jo Allen Gause, *Senior Program Officer, Capacity*
Rosalind Gomes, *Accounting/Financial Assistant*
Abdelmename Hedhli, *Visiting Professional*
James Hedlund, *Special Consultant, Safety Coordination*
Alyssa Hernandez, *Reports Coordinator*
Ralph Hessian, *Special Consultant, Capacity and Reliability*
Andy Horosko, *Special Consultant, Safety Field Data Collection*
William Hyman, *Senior Program Officer, Reliability*
Michael Marazzi, *Senior Editorial Assistant*
Linda Mason, *Communications Officer*
Reena Mathews, *Senior Program Officer, Capacity and Reliability*
Matthew Miller, *Program Officer, Capacity and Reliability*
Michael Miller, *Senior Program Assistant, Capacity and Reliability*
David Plazak, *Senior Program Officer, Capacity*
Onno Tool, *Visiting Professional*
Dean Trackman, *Managing Editor*
Connie Woldu, *Administrative Coordinator*
Patrick Zelinski, *Communications/Media Associate*

ACKNOWLEDGMENTS

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program, which is administered by the Transportation Research Board of the National Academies. The project was managed by Steve Andrle, Deputy Director, SHRP 2.

The Parsons Brinckerhoff research team acknowledges Peter Vovsha (Principal Investigator 1) and Bob Donnelly (Project Manager) of Parsons Brinckerhoff; Mark Bradley (Principal Investigator 2) and John Bowman of Mark Bradley Research & Consulting; Hani Mahmassani (Principal Investigator 3) of Northwestern University; Tom Adler of Resource System Group; Kenneth Small and David Brownstone of the University of California at Irvine; Kara Kockelman of the University of Texas at Austin; Jean Wolf of GeoStats; and Frank Koppelman.

FOREWORD

Stephen J. Andrle, *SHRP 2 Deputy Director*

Driver response to congestion and road pricing is an essential element to forecasting the future use of roadway systems and estimating the effect that pricing has on demand and route choice. Though many studies have been conducted in the past and revenue studies are routinely done for proposed toll roads, there is still a need for improving the behavioral basis for forecast. The objective of this project was to develop mathematical descriptions of the full range of highway user behavioral responses to congestion, travel time reliability, and pricing. These descriptions were achieved by mining existing data sets. The report estimates a series of nine utility equations, progressively adding variables of interest. This research explores the effect on demand and route choice of demographic characteristics, car occupancy, value of travel time, value of travel time reliability, situational variability, and an observed toll aversion bias. The primary audience for this research is professionals who develop travel demand and traffic forecasts. Policy makers may also have an interest in the behavioral findings that could have policy implications. Equations for commercial drivers were not developed since their routes are normally determined, in part, by contracts and company policies.

The researchers for this study identified both revealed and stated preference data sets that could be mined to estimate equations on driver responses to congestion and tolls. The primary data sets were from Seattle and New York. Supporting data sets, used for testing transferability of the equations, included San Francisco, Minneapolis, Chicago, San Diego, Orange County (CA), and Baltimore. A hierarchical choice framework was used. The choice framework considers first residential location and activities, followed by primary destination and intermediate stops, mode of travel, occupancy (when applicable), time of day, departure window, and finally route choice.

The basic utility equation features travel time and cost with coefficients estimated from the data sets. Additional levels of disaggregation may be used depending on the availability of data. In the next level, the equation specifies time to mean "free flow" and "congested" time. The data analysis indicates that drivers perceive every minute driving in congested conditions at 1.5 to 2.0 times longer than free flow travel time. In the next level, which adjusts the cost term for income, research shows that the value of travel time increases with income, but not linearly. The cost term is subsequently disaggregated by auto occupancy, followed by personal characteristics such as trip purpose, age, and gender. Sensitivity testing shows that segmentation by trip purpose is significant, but other personal characteristics are not extremely significant. Travel time reliability, considered in the next level, is the standard deviation of travel time adjusted for distance. This equation recognizes that the value of travel time reliability for short trips (e.g., 5 miles), especially trips to and from work, is greater. The next variable revealed from the data is a toll aversion bias, representing a psychological perception over and above time-cost trade-offs. The toll aversion bias is equivalent to 15–20 minutes of travel time even in areas with a long history of toll roads. The final term in the complete equation represents unobserved heterogeneity. This variable is significant because it represents what may be called "trip pressure" or other situational

factors in which there is a penalty for lateness (e.g., trips to the airport or to pick up children). People making such trips are often willing to pay a toll rate higher than demographic or trip purpose characteristics would indicate.

This research reveals a number of policy implications. Drivers place a value on travel time across a wide range from $5 to over $50 per hour and approaching $100 per hour when trip pressure is high. Therefore, toll levels have to be significant to influence congestion. Travelers' responses to congestion and pricing are also dependent on the options available. Driver response to congestion and pricing usually escalates from changing a route or departure time, to switching to transit if available, to rescheduling trips, and finally moving or changing jobs. Providing travel options is an important complement to a road pricing strategy that is aimed at reducing congestion. Finally, improvements to travel time reliability are as important as improvements to average travel time. This implies that operational improvements and information provided to travelers may be as valuable as increases in speed.

The report contains extensive documentation on the estimation of these models and the policy implications. It also contains insights on the value of travel time reliability and the use of reliability in travel demand and simulation models.

CONTENTS

Executive Summary

Organization

Project SHRP 2 C04, Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand, reviewed and advanced the state of the practice in modeling the effects of highway congestion and highway pricing on travelers' decisions, including choices of facility, route, mode, and time of day (TOD). This Executive Summary is intended for those who do not have extensive experience in travel demand modeling, but do wish to learn about and apply the results of the modeling research. The Executive Summary summarizes the objectives, data, methods, and key findings of the research. Each finding is accompanied by a discussion of the behavioral modeling issues, analysis results, and implications for transportation policy.

Reader Navigation

The C04 research aimed at breakthrough advances in travel demand modeling, and simulation was necessarily conducted within a highly specialized conceptual, mathematical, and technical context. Many important aspects of the work can only fully and meaningfully be described with theoretical constructs and important technical details that are difficult to comprehend by most general readers and even many model practitioners. The project team addressed this dilemma by maintaining continuity in the discussion, which can include both the more accessible content and the highly technical content, and offering navigation guides to the reader. Each section of the report typically begins with a discussion aimed at all readers; highly technical details are placed in Appendix A or are flagged in Chapter 6 with the header Related Technical Detail.

Technical Report

The Technical Report is intended for practitioners who perform or direct modeling of highway congestion and pricing policies and for planners and decision makers who wish to gain a deeper technical appreciation and understanding of particular issues. The discussion is organized around specific modeling questions and hypotheses that were tested, rather than around specific data sets and choice models.

Chapter 1 describes the research objectives and methodology. Chapter 2 describes the review and selection of data sources for tests of the travel demand models. The main research results described in Chapter 3 are organized by model types and features, and the model components are compared back-to-back. Chapter 4 covers the network simulation procedures for congestion and pricing studies, and Chapter 5 describes incorporation of the results into

practice. Conclusions, main findings, and recommendations for future research are summarized in Chapter 6.

Technical Appendix and Supplemental Material

Appendix A comprises all technical details related to the specification and estimation of C04 models. An unabridged, unedited version of Chapter 3 is available online at: www.trb.org/Main/Blurbs/168141.aspx. Appendix A provides the full statistical results for the key models estimated and discussed in Chapter 3. It is organized according to the types of models estimated and the data sets used. This appendix is intended for those with experience in estimating discrete choice models and will be useful to those who wish to estimate choice models on their own data sets. Full model specifications are included in tabular form, and models are numbered in a logical way corresponding to the discussion in Chapter 3.

Appendix A also consists of three short technical memoranda that document and detail the processing and development of data sets and network simulation formulations, algorithms, and calibration.

Summary of Objectives and Methodological Principles

Primary Objectives and Focus of Research

The C04 project was designed to (1) synthesize research findings from the past 30 years on travelers' responses to changes in both traffic congestion and the price of travel, (2) select the most important and well-founded behavioral hypotheses, (3) test those hypotheses statistically on the most suitable data sets available in the United States, and (4) identify ways in which the developed functions could be incorporated in operational models of travel demand and network simulations. The scope for the project and the range and quantity of statistical analysis performed were extensive. This summary report is a distillation of the issues studied and the most important findings.

The focus of the research was to identify the most important contextual influences on individual behavioral sensitivity to highway congestion and pricing and to provide guidance on the relative magnitudes of those influences. In practice, behavioral sensitivities are often expressed as elasticity—the percentage change in a behavioral outcome divided by the percentage change in travel time or cost. For example, the elasticity of vehicle miles traveled (VMT) in response to gasoline price has sometimes been measured at –0.2, meaning that a 10% increase in fuel price will lead to a 2% decrease in VMT. Although it is tempting to think that behavioral responses can be predicted by using such a simple measure, the reality is much more complex. For example, the elasticity of VMT with respect to price is also influenced by the income level of the traveler, the presence or absence of good transit alternatives, and opportunities such as buying a more efficient vehicle or finding a job closer to home.

The following sections summarize the methodological principles of the research.

Highway Utility (Generalized Cost) Formulation

The most common approach to dealing with context specificity is to use disaggregate discrete choice models to predict the choices that a given type of traveler will make for a given type of trip under specific circumstances. A discrete choice model assumes that when all important aspects of a choice are weighed, each choice alternative will have a resulting utility, or attractiveness, for each specific traveler. The probability that the traveler will choose a specific alternative is a function of that alternative's utility relative to the utility of all other available alternatives.

A typical formulation of the utility (U) of a highway alternative can be written as a linear function of trip time and cost (including all forms of pricing) in the following way:

$$U = \Delta + a \times Time + b \times Cost + \varepsilon \quad \text{(ES.1)}$$

where

Δ = alternative-specific constant;
a = auto time coefficient;
b = auto cost coefficient; and
ε = term that captures the residual effects of all variables that are not explicitly represented.

Note that Δ, a, b, and ε may be either single terms or functions of other variables (e.g., income or trip purpose). In certain simple choice contexts like route choice, in which all alternatives are qualitatively similar, the alternative-specific constants can all be assumed to be equal to zero and dropped from the equation. However, in more general choice contexts that involve different modes of travel or different TOD periods, it is essential to account for qualitative differences between alternatives. In these choice contexts, Δ may include additional explanatory variables beyond travel time and cost.

Because utility has no physical dimensions, the coefficients are generally interpreted as relative to each other, as ratios. For example, if the estimated coefficient for auto travel time is –0.02/minute, and the estimated coefficient for auto cost is –0.10/$1, then the ratio of the two coefficients (a/b) is equal to $0.20/minute, or $12/hour. This particular ratio (here designated as value of time [VOT]) can be interpreted as the additional price that a traveler would be willing to pay for a marginal decrease in travel time (or to avoid a marginal increase in travel time).

The willingness to pay for reduced travel time varies substantially depending on the characteristics of the traveler and the context of his or her particular trip. Typical U.S. modeling analyses have only accounted for a limited amount of this variation, usually segmenting by trip purpose and sometimes by income segments.

Key Behavioral Hypotheses

Most of the tested hypotheses were related to how travel time and cost enter the highway utility formulation. The simplistic linear form of Equation ES.1 was questioned in many respects. Its primary drawbacks relate to the unrealistic assumption of a constant VOT.

Previous research identifies three major aspects of highway driving time that influence behavior and are perceived as important components of highway level of service (LOS): quantity (duration of time in the vehicle); quality (amount of stress or pleasure caused by the particular driving conditions); and reliability (level of uncertainty with respect to travel time and congestion levels). Each of these three travel time aspects and travel cost are likely to influence travel choices differently, although the corresponding effects are often intertwined.

Travel Time Quantity

The duration of travel time influences travelers' schedules and the alternative uses of the time they spend traveling. People with busy schedules are usually more likely to seek an alternative that will gain them a shorter-duration trip. For frequent trips (e.g., commuting), people may be more aware of the time duration difference between different travel alternatives and also more able to change their activity schedules to make optimal use of the time saved. Also, the longer is the block of time saved by choosing a specific travel option, the more likely it is that a traveler will perceive the time saved. The project team hypothesized that the utility of travel time duration is likely to vary according to the time constraints and activity schedule of the traveler and the frequency, TOD, and distance of a specific trip.

In practice, engineers often do not have accurate data on travelers' time constraints or trip frequency, particularly for forecasts. Trip purpose serves as a proxy, as regular trips to work or school tend to be the most frequent trips, and are often made by those with the busiest schedules.

Travel Time Quality

Many drivers find driving in stop-and-go traffic more stressful than driving in free-flow conditions and would be willing to pay more to avoid time spent driving in heavy traffic. For example, many drivers will drive a longer distance to avoid congestion bottlenecks; they may spend the same amount of time or more in traveling, but in less stressful conditions. These differences in congestion levels tend to be related to differences in travel time reliability (discussed below), but those reliability effects are greater than the effects referred to here, which relate only to the physical and mental stress of time spent in the vehicle, and which would exist even if the duration of travel time was completely predictable.

Although the stress caused by driving in different conditions may vary from driver to driver, the analyst can expect to find some systematic effects. In general, people place a higher value on time savings that arise from reductions in congestion levels than on time savings from other types of system changes, such as the introduction of shorter-distance routes or closer destinations. This further implies that time savings will be valued more for specific facilities and times of day when congestion levels are the highest. People often report that driving becomes more stressful as the duration of the commute trip increases beyond 30 or 40 minutes. Thus, one might expect the value of time savings to increase with trip distance or duration, but not in a linear fashion. A possible explanation of the trip-length effects lies in the structure of the entire daily activity pattern rather than in the commute trip itself. When commuting time grows beyond 2.5–3 hours per day and is combined with 8.5–9 hours at work, the total work–commute time makes it difficult to incorporate other activities of a significant duration.

Travel Time Reliability

Even when average travel times for two highway alternatives are the same, drivers will generally prefer the more reliable alternative (least day-to-day variability in travel time) or the lower risk that the travel time will be significantly longer than average. Qualitative and quantitative research has indicated at least three following reasons for this preference:

1. Negative Consequences of Arriving Late at One's Destination. These can include missing an appointment, missing a flight, or losing pay for work time. This consideration gave rise to the *schedule delay* concept in measuring travel time reliability, which is discussed in the Technical Report.
2. Need for Buffer Time to Avoid Arriving Late. Travelers concerned about a late arrival must begin their trip earlier than they would if the travel time were more reliable. This behavior will avoid most instances of arriving late, but at the expense of departing earlier and sometimes arriving too early. This consideration gave rise to a concept in measuring travel time reliability that operates with estimates of buffer time based on the travel time distribution shape.
3. Discomfort Related to the Uncertainty of How Long the Trip Will Take on any Given Day. This approach operates with the simplest quantitative measures (e.g., standard deviation) of travel time variability and probably has the best chance of being incorporated in operational models.

Quantitative research into VOT variability and value of reliability (VOR) has lagged because of the lack of data on the day-to-day travel time variability that drivers face for particular trips. Measuring variability requires an estimate of the travel time distribution, which

can be translated into measures such as standard deviation or buffer time represented by the 80th or 90th percentile of travel time versus the median. Estimates of the distribution for an entire trip distance from origin to destination (O-D) are also needed. However, because the correlation between travel times on different links in a network is a complex function of the network configuration and structure of the traffic flows, the distribution for an O-D time is not simply the sum of the distributions across highway links. As a result, O-D–level estimates are quite rare, unless one has extensive global positioning system (GPS)–based trace data over time, or some other means of estimating O-D variability (such as one specifically applied in the current study). Most previous quantitative research into VOR has been based solely on stated preference (SP) data, particularly in Europe. This project makes one of the most thorough efforts to date to overcome these data limitations and obtain estimates using revealed preference (RP) data from different regions of the United States.

Travel Cost

Various auto-related costs can influence the utility of a particular trip alternative (in terms of route, departure time, carpooling, and so forth). Although costs such as vehicle maintenance and insurance can vary with the mileage driven, the travel decisions made for a given trip are more strongly influenced by the direct costs of parking, tolls, and fuel. This research focuses on toll costs because they are most important in terms of future road-pricing policy and because they tend to provide the clearest and most statistically advantageous contexts for measuring the importance of price in choosing between highway travel options.

The sensitivity of travel behavior to a specific type of travel cost depends primarily on (1) how much of the cost a traveler actually has to pay (e.g., cotravelers may share the cost of fuel or tolls) and (2) how affordable that cost is for the particular traveler. The team hypothesized that the most important contextual differences determining the sensitivity of travel behavior to price are related to income, vehicle occupancy, and travel purpose. This behavioral mechanism should be considered in the regional network context. In particular, the presence of a reasonable transit alternative plays a major role in determining the final outcome of congestion and pricing.

Data and Methods to Test Behavioral Hypotheses

Only a few of the more than 100 surveys identified were adopted. These were chosen for the probability that they could support various travel time reliability measures, as well as toll and nontoll routes for the same trip. The team relied on the three types of travel survey data described below.

Revealed Preference Data

RP data are observed data on actual choices made by travelers. Although RP data are always preferable, the existing RP data sets have many limitations. RP data are collected from travel diaries in which survey respondents report all of the travel and activities they undertake in the course of a representative weekday. Because the respondents typically do not report either all of the available (but nonchosen) trip alternatives, or all of the travel times and costs related to the chosen and nonchosen alternatives, those supply-side LOS measures must be inferred from representations of the highway and transit networks. This task can be expensive and time-consuming. Only those RP data sets that can be supported by a well-calibrated regional travel demand model with network simulations implemented for multiple periods of a day are usable. The RP-type Household Surveys in New York City and Seattle, Washington, were adopted for many statistical tests because they could be supported by the necessary LOS variables.

Stated Preference Data

SP data are responses by survey respondents to questions about hypothetical travel situations. These data are collected in SP choice experiments that are customized around the context of an actual reported trip that a respondent has recently made. The SP approach can be used to study the demand for an alternative that does not actually exist, such as a new tolled highway facility. The choice set of available travel alternatives and the related times and costs are explicitly specified for respondents, which avoids the difficulty and expense of inferring those supply variables after the fact. The obvious disadvantage of SP data is that the choices are hypothetical, so there is less confidence that the analysis results will reflect true behavioral relationships. This concern is perhaps strongest in the case of complex variables such as travel time variability, which are difficult to portray clearly to SP respondents. In the current project, SP data were a complementary source that helped explore situations that were not observed in RP surveys.

Experimental Revealed Preference Data

The approach of experimental revealed preference data merges the best characteristics of RP and SP approaches by measuring actual choices from experimental contexts created on actual travel networks. An example is a system of tolls introduced temporarily in a road network and charged via electronic tolling, but only to specific drivers who are participating in the experiment. Experimental RP data from the recent Traffic Choices Study in Seattle were used in the current research.

Data Analysis Methods and Key Data Sets

After a careful assessment of RP and SP data sets from across the United States, the team selected a handful of data sets that would best support the planned range of analyses (Table ES.1). The New York and Seattle metropolitan regions were selected as the primary regions for RP data analysis because the Household Travel Survey data sets from these areas, together with associated highway and transit network supply data, could support detailed disaggregate model estimation. In addition, the Puget Sound Regional Council had carried out an innovative mileage-based tolling experiment that provided a unique source of RP data that complemented the other Seattle region RP and SP data sets. SP data sets from San Francisco and Los Angeles, California, were also analyzed.

Table ES.1. Data Sets Used for Analysis

Geographic Area	Planning Agency	RP Data	SP Data	Experimental RP Data
New York metropolitan region	New York Metropolitan Transportation Council	1997 Household Travel Survey (1-day diary)		
Seattle metropolitan region	Puget Sound Regional Council	2006 Household Travel Survey (2-day diary)	2006 highway toll SP experiment (follow-up with 2006 survey respondents)	2006 Travel Choices tolling experiment (GPS-based tolling simulation)
San Francisco County	San Francisco County Transportation Authority		2007 downtown area pricing SP experiment	
Los Angeles County	Los Angeles Metropolitan Transit Authority		2008 managed toll-lane SP experiment	

Table ES.2. Types of Choice Models Supported by the Data Sets

Data Set	Toll versus Nontoll Route Choice	Departure Time Choice	Mode and Occupancy Choice	Other Choice Dimensions
New York region 1997 RP household survey	X	X	X	X
Seattle region 2006 RP household survey		X	X	X
Seattle region 2006 Traffic Choices pricing experiment	X	X		
Seattle 2006 SP toll experiment	X	X		
San Francisco 2007 SP area pricing experiment		X	X	
Los Angeles 2008 SP managed-lane experiment	X	X	X	

Table ES.2 summarizes the aspects of choice represented in each of the primary data sets. The New York regional household travel survey RP data set supports the widest range of modeling. Because this region has tolled bridges and tunnels and has collected data on whether those tolls were paid on the respondents' actual trips, the data can be used to model the choice between tolled and free routes. The data also support models of departure time choice (analyzing the times of day that respondents chose to make their trips as a function of congestion levels) and mode choice (the decision to drive alone, carpool, or use bus or rail as a function of the travel times and costs of the different modes). New York, a transit-rich area with an extreme level of congestion and well-established toll facilities, provides a variety of data for exploring trade-offs between travelers' travel time and costs.

Because travel diary data sets provide a complete picture of a representative day's travel, they can be used to model other dimensions of travel choice, such as number of trips made for different purposes, as well as longer-term decisions, such as the number of automobiles to own.

The Seattle region 2006 household RP survey data are comparable to those from the New York region, with the exception that the Seattle region had few tolled facilities at that time. Thus, the Seattle data will not support models of toll versus nontoll route choice, although they can be used to model the influence of other travel costs such as fuel and transit fares on mode choice. In contrast, the Seattle Traffic Choices experimental RP data offered tolls that varied by distance, facility type, and TOD, so these data are suitable for modeling route and departure time choices. The GPS-based data collection method did not provide information on vehicle occupancy or the use of non-auto modes, so its data are not useful for modeling mode or auto occupancy choice or other choice dimensions.

Each of the analyzed SP data sets focused on specific choice dimensions with respect to tolls. All of them offered respondents toll levels that varied by TOD, so the data sets are useful for modeling the effects of pricing (and congestion) on departure time. Some also offered the options of free alternative routes or competing transit alternatives, so those data sets can be used for route choice or mode choice analysis, respectively. There were at least two RP data sets and two SP data sets to model each of the main travel choice dimensions.

In order to focus on the national goals established for SHRP 2 and because of the necessity of managing the data sets and supporting them by the regional travel models and network simulations available to the project team, data from other countries were not considered. The team believes most of the results can be extended (at least qualitatively or conceptually) to areas outside the United States. But because regional conditions play a significant role in shaping travel behavior, a direct transfer of model structures and coefficient values to areas outside the United States is not recommended.

Analysis Approach for Improved Demand Modeling

The team tested the same types of variables and functional specifications of the generalized cost on multiple data sets and choice contexts and looked for consistencies that suggested the most reliable practices and productive paths for modeling the effects of highway congestion and pricing. This systematic approach involved the following steps:

- The first models used the most basic specifications for each data set and each choice dimension (e.g., route or mode choice) separately, using only simple time and cost variables included in Equation ES.1;
- Additional variables (e.g., segmentation variables for income and car occupancy) were then systematically added to the models;
- Model complexity was gradually increased when possible to simultaneously model different travel decisions using nested hierarchical choice structures, such as a joint model of departure time choice and mode choice; and
- The probabilistic distribution (spread) of the key travel time coefficients was estimated using advanced mixed logit specifications.

This stepwise analysis approach across several data sets revealed the stability and generality of the modeling results and allowed a recommendation for a generalized cost specification.

The team recommends using a highway utility function of the general form (Equation ES.2), which is an extension of the simplified form (Equation ES.1):

$$U = \Delta + a_1 \times Time \times (1 + a_2 \times D + a_3 \times D^2) + b \times [Cost/(I^e \times O^f)] + c \times STD/D \qquad \text{(ES.2)}$$

where

Δ = alternative-specific bias constant for tolled facilities;
a_1 = basic travel time coefficient (ideally estimated as a random coefficient to capture unobserved user heterogeneity);
$Time$ = average travel time;
a_2, a_3 = coefficients reflecting the impact of travel distance on the perception of travel time;
D = travel distance;
$Cost$ = monetary cost (e.g., tolls, parking, and fuel);
I = (household) income of the traveler;
O = vehicle occupancy;
e, f = coefficients reflecting effect of income and occupancy on the perception of cost;
STD = day-to-day standard deviation of the travel time; and
c = coefficients reflecting the impact of travel time (un)reliability.

Equation ES.2 includes travel cost explicitly scaled by income and vehicle occupancy; travel time reliability (variability) is included separately from the typical, or median, travel time. The team found that the standard deviation of travel time divided by distance gave the strongest and most consistent results for the reliability effect. Conceptually, this variable represents the day-to-day variability in highway travel speed (time divided by distance is the inverse of speed).

The distance-based term by which travel time is multiplied expresses an important distance effect on travel time perception and VOT. This term represents a polynomial function of distance that scales travel time in the following way:

$$Time \times (1 + a_2 \times D + a_3 \times D^2) \qquad \text{(ES.3)}$$

For travel segments with a short average distance, the distance-related effects are insignificant, and the entire multiplier can be dropped. For longer segments, such as commuting to work in large metropolitan areas, the distance effects are significant. VOT can grow or decline with distance depending on the sign and magnitude of coefficients a_2, a_3.

The suggested form for accounting for travelers' perceptions of travel costs corresponds to the monetary cost scaled by power functions of both income and vehicle occupancy:

$$Cost/(I^e \times O^f) \tag{ES.4}$$

Coefficients *e, f* lie in the unit interval. If the coefficient is close to zero, the corresponding scaling effect becomes insignificant and can be dropped. If the coefficient is close to 1.0, the scaling effect reaches the maximum. The most statistically significant values were found by exploring the entire range of possible values.

The recommended and most statistically significant main measure of travel time reliability is the day-to-day standard deviation of travel time by auto, divided by distance:

$$STD/D \tag{ES.5}$$

This reliability measure is especially practical for RP-based models because it obviates one of the most problematic features of most RP data sets: correlation between travel time, travel cost, and practically any travel reliability measure, including standard deviation or buffer time. This measure has a plausible behavioral interpretation: Travelers may perceive travel time variability as a relative (qualitative) measure rather than an absolute (quantitative) measure. This behavioral assumption is appealing in the context of the entire highway utility (Equation ES.2), in which travel time and cost are included in an absolute fashion; thus the reliability term plays a complementary role and explains what has not yet been explained by the time and cost terms.

VOT can be calculated as

$$VOT = \frac{a_1}{b} \times (1 + a_2 \times D + a_3 \times D^2) \times (I^e \times O^f) \tag{ES.6}$$

In a general case, VOT is a function of travel distance, income, and car occupancy for each travel segment. If the model is explicitly segmented by these variables, then the formula for VOT can be simplified and made specific to each segment by differentiation of the basic time and cost coefficients.

VOR can be calculated as follows:

$$VOR = \frac{c}{b} \times \frac{(I^e \times O^f)}{D} \tag{ES.7}$$

Like VOT, VOR is a function of travel distance, income, and car occupancy for each travel segment unless a more detailed explicit segmentation is applied. VOR is inversely proportional to distance (i.e., the longer the distance, the greater the magnitude of the reliability measure), although a longer distance tends to dampen travel time variation. The portion of travel time variability that is proportional to the average travel time is accounted for in the loaded travel time coefficient a_1. Thus, only the residual variation of travel time expressed as standard deviation per unit distance is accounted for in the reliability term.

Finally, the reliability ratio (*RR*) can be calculated as a measure of the relative importance of reduction of (un)reliability versus average travel time savings:

$$RR = \frac{VOR}{VOT} = \frac{c}{a_1} \times \frac{1}{(1 + a_2 \times D + a_3 \times D^2) \times D} \tag{ES.8}$$

Logically, the reliability ratio is a function of travel distance rather than a fixed value. The reliability ratio usually declines with distance, a fact that cannot be taken out of the functional form (Equation ES.2), in which the terms for (loaded) travel time and travel time reliability complement each other.

This form of highway utility (Equation ES.2), with its choices for route type, mode, car occupancy, and TOD, can be incorporated in operational travel demand models in the near future.

Through these primary choice dimensions, the impacts of congestion and pricing can be further propagated through the model system chain to affect destination choice, trip frequency, and other choice dimensions.

Impacts of Congestion and Pricing on Travel Demand: Behavioral Insights and Implications for Policy and Modeling

Key study findings are related to the estimated values of the parameters in Equation ES.2. These findings are presented as a series of 11 key behavioral insights, along with the implications of these insights for road pricing congestion management and improved travel modeling:

- Variation in VOT across highway users;
- Income and willingness to pay;
- Auto occupancy or group travel and willingness to pay;
- Constraints on TOD shifting (carpools and single-occupant vehicles);
- Importance of VOR and its relationship to VOT;
- Effect of travel distance on VOT and VOR;
- Evidence of negative toll bias;
- Hierarchy of likely responses to change in tolls and congestion;
- Summary of user segmentation factors;
- Avoiding simplistic approaches to forecasting; and
- Data limitations and GPS-based data collection methods.

Variation in Value of Time Across Highway Users

Key Finding

VOT varies widely, from $5 to $50/hour across income groups, vehicle occupancies, and travel purposes. There is significant situational variation (unobserved heterogeneity), with some people willing to pay almost nothing to save time, and with others willing to pay more than $100/hour.

Implications for Policy

The wide distribution of willingness to pay confirms that pricing can effectively serve the important function of market discrimination and demand management. Because most travelers have a relatively low willingness to pay, any price that affects all travelers, such as a general toll for all lanes of a highway, may influence demand at fairly modest levels. In contrast, prices for high-occupancy toll (HOT) and express lanes can be set at fairly high levels and adjusted to attract a relatively small percentage of travelers with the highest willingness to pay. Pricing policies should be applied after a careful analysis of possible negative implications for low-income users.

Implications for Modeling

Most models used for travel demand forecasting have assumed a single VOT. Only occasionally have different cost coefficients been used for different income groups and vehicle occupancy levels. Differentiation of VOT is even less typical in network simulation procedures. These practices result in significant aggregation biases that affect the accuracy of traffic and revenue forecasts. Whenever possible, random coefficients should be used to estimate the distribution of VOT across the population. For general use, newer activity-based forecasting models that use a microsimulation approach can simulate a different VOT for each person and trip, providing the most disaggregate treatment of VOT, and thus avoiding one important source of possible errors and biases in the forecasts.

Income and Willingness to Pay

Key Finding

Household and personal income has a strong relationship with VOT and willingness to pay, but the relationship appears to be less then linear. To account for the income effect, cost variables in travel models (including tolls) should be divided by household income, raised to a power in the range 0.6 to 0.8 depending on the trip purpose (e.g., for a power of 0.7, doubling income increases VOT by 62%; halving income decreases VOT by 38%).

Implications for Policy

The income effect is strong, so that many of the benefits of pricing are purchased by those who can most afford them, and equity considerations cannot be discounted. Lower-income travelers also derive benefits in the form of increased options, as well as improvements in traffic conditions if total capacity can be increased through priced facilities. The parallel effect of car occupancy mitigates the income effect. Low-income commuters have more opportunities to carpool and share commuting costs than do high-income commuters. High-occupancy vehicle (HOV) and HOT lanes, as well as transit, represent viable alternatives for low-income travelers.

Implications for Modeling

Forecasting models typically use income either in a simplified linear form to scale travel costs or as a segmentation variable, with different cost coefficients in different income ranges. Neither approach seems entirely appropriate. The assumption of linearity with income seems too strong, particularly in higher income ranges, and the piecewise linear approach often results in strong nonlinearities or discontinuities in the effect of income that do not have a strong statistical or behavioral basis. The recommended approach is empirically justified across a wide body of evidence and provides a smooth response surface for forecasting.

Auto Occupancy or Group Travel and Willingness to Pay

Key Finding

Auto occupancy has a strong estimated relationship with VOT and willingness to pay. The relationship appears to be slightly less than linear. To account for occupancy effects, cost and toll variables in travel models should be divided by occupancy raised to a power in the range 0.7 to 0.8.

Implications for Policy

Because a group of vehicle occupants is generally willing to pay more than a solo driver, a tolled facility is likely to attract a higher percentage of multioccupant vehicles than will a free facility, even if no special carpool discount is offered; in effect, a carpool discount is being offered to a group that tends to value it the least. However, ridesharing increases system capacity, and the conversion of HOV lanes to HOT lanes may potentially discourage carpooling among individuals with higher VOT by offering solo drivers the same travel time advantage without the added inconvenience of ridesharing. Thus, free or discounted toll lanes for carpools will encourage carpoolers, even if it does not attract much additional ridesharing.

In general, low-income commuters have a higher probability of forming a carpool than do high-income commuters. Low-income workers normally have a fixed work schedule, which simplifies carpooling logistics, and they tend to live in dense residential clusters where the process of collecting and distributing passengers requires minimal extra time. In addition, low-income jobs tend to form clusters of multiple jobs. This factor may vary with the structure of

the metropolitan area, however, as large clusters of high-income jobs may be present in the central business district (CBD).

The higher opportunity for carpooling among low-income workers mitigates the equity concerns regarding pricing, because costs can be shared within the carpool. Faced with significant tolls, high-income workers can only switch to transit, but low-income workers can use transit or HOV lanes. This consideration is frequently missed in policy analysis of pricing projects, which may result in exaggerated equity concerns.

Implications for Modeling

Dividing travel cost by vehicle occupancy is a fairly standard practice in applied modeling. The team's main recommendation is to divide costs by a function of occupancy that is somewhat less than linear. Income-specific components in the car occupancy choice are needed to reflect differential opportunities to carpool by income. Simplified approaches based on average occupancy coefficients tend to mask these important effects and portray pricing projects in an extreme way with respect to different income groups.

Constraints on Time-of-Day Shifting: Carpools and Single-Occupant Vehicles

Key Finding

Although commute carpools generally have a higher VOT, they also tend to have tighter scheduling constraints and tend to be less flexible in their capacity to shift departure time away from the peak period and hour.

The team consistently found that ridesharing commuters are more likely to travel in the heart of the peak periods than those who drive alone. Carpool commuters must coordinate their commute schedules with cotravelers, so it is less likely that they can adjust their departure times earlier or later to avoid peak congestion or pricing.

Implications for Policy

Because carpooling commuters generally have little opportunity to retime their trips to avoid peak congestion times, TOD pricing and other peak-spreading policies will tend to be less successful in influencing their behavior. To avoid inadvertently discouraging ridesharing, policies should be designed to offer some level of advantage of travel time or price (or both) for HOVs. Congestion pricing policies could be more effective if they were accompanied by policies encouraging employers in the CBD (or other relevant congestion pricing zones) to shift working hours and introduce flexible or compressed work weeks.

Implications for Modeling

Modeling studies intended to predict peak spreading behavior and response to TOD pricing should include different sensitivities for different car-occupancy levels. In general, the propensity to switch from the peak hour to a different hour is inversely proportional to vehicle occupancy.

Importance of Value of Reliability and Relationship to Value of Time

Key Finding

Travelers tend to value variation in travel time reliability (day-to-day variability) at least as highly as they value variations in the usual travel time. Various ways of specifying the variability variable

were tested, but the measure that produced the most consistent results was the standard deviation of travel time divided by journey distance.

Evaluating the estimation results to impute the reliability ratio (the value of reducing the standard deviation of travel time by 1 minute divided by the value of reducing the average travel time by 1 minute, or VOR/VOT for an average trip distance) obtained ratios in the range 0.7 to 1.5 for various model specifications. SP studies from Europe give typical values in that same range for auto travel, with higher ranges up to 2.5 for rail and transit travel. The SP results, however, indicate that the estimates may vary depending on how the reliability concept is presented to respondents. Thus, it is crucial to obtain new estimates based on actual choices at the trip level.

Implications for Policy

Highway investments that can improve travel time reliability will tend to be just as beneficial for travelers as investments to reduce typical travel times. This finding underlines the importance of addressing key bottleneck points and using transportation systems management and intelligent transportation systems to monitor and adapt to congestion levels on the network, as well as systems to avoid nonrecurrent congestion and to recover from it as quickly as possible. For managed lanes and other priced facilities, the "guarantee" of a reliable travel time may be of great value. This makes variable pricing, and especially that of dynamically priced lanes, one of the more effective pricing forms that are attractive for the user. These findings also emphasize the importance of effective accident management, as the consequences of traffic accidents constitute a significant share of long delays.

Implications for Modeling

Although models can be estimated using measures of day-to-day travel time variables from real and simulated highway networks, further progress is needed before this method is feasible for most travel demand forecasts, particularly in terms of widespread collection of data for actual levels of travel time variability at the O-D level. Certain technical issues must also be resolved on the network simulation side, specifically the incorporation of travel time reliability in route choice and the generation of O-D travel time distributions instead of average travel times. In the near term, this method may be most applicable to corridor- and facility-level forecasts. Some simplified implicit measures of reliability (such as perceived highway time by congestion levels, as explained below) can be applied with the existing model structures and network simulation procedures.

Effect of Travel Distance on Value of Time and Value of Reliability

Key Finding

Savings on average or typical travel time (VOT) are valued more highly for longer trips than for short trips, except for a special effect on commuting trips over 40 miles. For VOR, there is a relative damping effect for longer trips. These findings suggest the efficacy of using higher-priced managed lanes to address key bottlenecks and lower distance-based tolls on the wider highway network.

Implications for Policy

Traffic bottlenecks tend to increase the variability (unreliability) of all trips that pass through them, regardless of total trip distance, and the results indicate that all travelers will derive considerable benefit from making the system more reliable. In contrast, improvements that increase average speeds or reduce travel distances without substantially improving reliability will not be valued very highly by those who only use the facility for a short distance. Thus, distance-based tolls are appropriate in general, but higher prices that are not based on distance may be more appropriate for addressing key bottlenecks.

Implications for Modeling

Because VOT and VOR tend to vary with O-D trip distance, using a constant VOT and VOR for a wide range of trip lengths is an unreasonable simplification pertinent to most travel models. For the most accurate predictions, this distinction should be used in demand-forecasting models.

Evidence of Negative Toll Bias

Key Finding

There is a significant negative bias against paying a toll, regardless of the toll amount. This preference is generally supported across travel purposes by RP and SP data, as well as by research in behavioral economics. The estimated toll penalty effect for auto trips is generally equivalent to as much as 15–20 minutes of travel time.

Implications for Policy

The resistance to paying a toll appears to present an obstacle to the effective widespread introduction of congestion pricing policies. However, a pricing policy can be effective even if only a limited proportion of drivers pay the toll, and as with VOT, the resistance to paying any toll at all may vary widely across the population. In that sense, toll bias becomes another dimension of market discrimination, similar to VOT. Resistance can be overcome by a guaranteed superior LOS in terms of travel time savings and improvements in reliability. Tolling existing facilities only to collect revenue, but without a substantial LOS improvement, would generally be perceived negatively by highway users.

Resistance to paying a toll is likely to fade as road pricing becomes more ubiquitous and more convenient. In the past, drivers had to wait in lines to pay tolls, which in itself could explain a good deal of resistance to tolls. Now, electronic tolling has made paying the toll both faster and less noticeable in terms of the amount of money being spent. The more widespread becomes electronic road pricing, the more it can be expected that antitoll bias will reduce.

Implications for Modeling

Antitoll thresholds are avoided in forecasting on the basis that they are not rational in economic terms. Empirically, however, they do appear to be real, so they should be included to obtain the most accurate results, at least for short-term forecasts. This bias will result in a more conservative traffic and revenue forecast if travel time savings are insignificant, but it also may result in a more optimistic forecast for pricing projects that improve travel time significantly. For longer-term forecasts, it may be appropriate to explore scenarios with reduced or eliminated antitoll bias terms.

Hierarchy of Likely Responses to Changes in Tolls and Congestion

Key Finding

Traveler responses to congestion and pricing depend on the range and attractiveness of available alternatives. The models estimated for this project covered a range of travel choices. When possible, nested hierarchical models were estimated to determine which types of choices are most sensitive to travel time and cost changes. The highest propensity for change appears to be between tolled and nontolled routes. A change of route requires little or no adjustment in travel schedule, and the choice can even be made en route. Travelers also show a fairly high propensity for making minor shifts in departure time of an hour or less, since the smaller is the shift, the less rescheduling of activities is required, and the more familiar the traveler is likely to be with the typical traffic conditions over time.

Somewhat less likely are changes in travel mode or car occupancy, which may include switching between auto and transit or between driving alone and ridesharing. Mode shifting is most prevalent for commute trips and other frequent trips for which information about transit services or possible carpoolers is most available or worth investigating.

Less likely responses to changes in congestion or pricing are changes in the choice of destination locations, the rescheduling of trips to very different times of day, and changes in the frequency of making trips from home. These types of changes are the least likely for activities that are most constrained, such as work and school trips or medical appointments. For more flexible types of trips, these types of shifts may actually be more likely than changing the mode of travel.

In the longer term, people may make greater changes as opportunities arise and life-cycle transitions occur. These shifts include changing the number or type of vehicles owned (or both) and the location of home, work, school, and other key travel anchor points relative to one another. The present project outlines an approach to modeling longer-term responses to congestion and pricing by means of accessibility measures that are derived from the estimated primary choice of route, mode, and TOD.

Implications for Policy

Decisions influencing traffic congestion and the cost of driving can affect travel behavior, and the relationships are often complex and can shift over time. This aspect of travel behavior argues for using advanced demand simulation models to guide policy, rather than relying on mental models and experience. The most predictable effects tend to be those that require only minor adjustments, such as choosing to travel at a slightly different TOD. To make pricing policies more effective in tackling congestion, the presence of competitive alternative modes and destinations should be carefully considered. Pricing policies are most effective in combination with transit improvement and smart land use development.

Implications for Modeling

Modeling systems should be able to represent the influences of travel time and cost on all of the types of decisions listed above, and the models should be integrated so that appropriate relative sensitivities are reflected at the different hierarchical levels. These relative sensitivities should also allow for variation in travel segments and travelers. Such modeling requires an activity-based microsimulation model, ideally used in combination with accurate dynamic simulation of traffic congestion.

Summary of User Segmentation Factors

Key Finding

Many potential factors can affect VOT, VOR, or traveler responses to congestion and pricing, including person, household, land use, and travel characteristics. It will never be possible in regional travel models designed for long-term forecasting to account for all the details of user characteristics. However, it is possible to account explicitly for the most important and systematic effects and to apply reasonable assumptions about the probabilistic distributions of VOT and VOR in order to account for the residual heterogeneity.

Implications for Policy

Most of the important factors that affect traveler responses to congestion and pricing are highly differentiated by highway user groups. In calculation of user benefits, the analysis must be implemented with a user segmentation that at a minimum includes trip purpose (work and nonwork); income group and car occupancy (three to four categories each); and commuting distance and household size (two to three categories each). It is highly desirable to account for significant unobserved user heterogeneity and situational variability by applying probabilistic VOT/VOR rather than deterministic VOT/VOR. Simplified methods that operate with an average VOT/VOR are subject to significant aggregation biases and will not adequately portray a pricing project.

Implications for Modeling

Segmentation is crucial for policy evaluation, and modeling systems should be segmented according to the main effects described above. Traditional four-step demand models and static traffic assignments, still the most common tools in practice, are of little use because limited segmentation is one of their major constraints. In addition, it is practically impossible to incorporate distributed parameters in these aggregate constructs. Activity-based models (ABMs) on the demand side and dynamic traffic assignment (DTA) on the network simulation side offer the potential for significantly better platforms for modeling highway congestion and pricing because they are both based on the concept of individual microsimulation.

Avoiding Simplistic Approaches to Forecasting

Key Finding

Although many key effects and tendencies related to the highway utility function are similar across data sets and regions in the United States, many additional effects associated with person types, household composition, transit availability, and land use are specific to each region. Therefore, any simplified surrogate equations or elasticity calculations need to be interpreted and applied with caution.

Implications for Policy

Interregional comparisons and analogies and general rules with respect to expected demand elasticity in relation to congestion and pricing must be applied cautiously. In general, they should not be used for evaluating pricing projects and policies or comparing different pricing alternatives. Properly portraying congestion and pricing effects, as well as the large magnitude of possible impacts (positive or negative), fully justifies a serious modeling approach with a corresponding data collection effort.

Implications for Modeling

The functional forms for the highway utility function developed in the present research should be applied within a framework of regional travel models in which all needed structural inputs and market segments can be supported. Such travel models can fully address regional specifics and take advantage of available data. The best framework is a complete regional travel model system in which an advanced travel demand model (preferably of the activity-based microsimulation type) is integrated with an advanced network simulation tool (preferably DTA with microsimulation of individual vehicles).

Data Limitations and Global Positioning System–Based Data Collection Methods

Key Finding

The availability of data sets adequate to support the analyses undertaken in this study was extremely limited, especially for travel time reliability. This kind of difficulty should decrease, because the use of GPS and probe vehicles and other distributed wireless technologies to collect data on actual travel times and speeds is growing rapidly.

Implications for Policy

With more comprehensive and credible data on travel times and speeds, including measures of travel time reliability, policy makers will have a significantly better basis for advocating new

projects and policies, including pricing. The entire issue of improving travel time reliability can finally shift from qualitative analysis to quantitative analysis.

Implications for Modeling

New sources of information are essential for estimation and calibration of travel demand models and network simulation tools. Crude LOS variables created by static assignment procedures have always formed one of the weakest components in travel modeling, frequently manifested in illogical values of model coefficients that must be constrained to ensure reasonable model sensitivities. All travel demand and network simulation models would benefit from better estimates of O-D travel times by TOD. Special benefits would be provided to and could be exploited by advanced models that incorporate travel time reliability measures.

Network Simulation Models to Support Congestion and Pricing Studies

Salient points of the C04 research with respect to network simulation tools include the following:

- Need for Microsimulation. Capturing user responses to pricing and reliability is best accomplished through microsimulation of individual traveler decisions in a network platform; a time-dependent analysis tool is required because the time dimension is essential to evaluating the impact of congestion pricing and related measures. Microsimulation of individual traveler choices provides the most general and scalable approach to evaluate the measures of interest in this study;
- Need for More Robust DTA. The current generation of available simulation-based DTA models only considers fixed, albeit time-varying, O-D trip patterns. Greater use and utility will result from integrating DTA with an activity-based demand model and incorporating user attributes, including systematic and random heterogeneity of user preferences;
- Improved Algorithms for Regional Scale Modeling. Finding equilibrium time-varying flows has been based on the relatively inefficient method of successive averages, a method that does not scale well for application to large metropolitan networks. New implementations of the method of successive averages and other algorithms that exploit the vehicle-based approach of simulation-based DTA have been demonstrated on large actual networks in this research effort;
- Traveler Heterogeneity. Incorporating heterogeneity of user preferences is an essential requirement for modeling user responses to pricing in both travel demand models and network simulation tools. New algorithms that exploit nonparametric multicriteria shortest-path procedures allow VOT to be continuously distributed across users. Efficient implementations of these algorithms have been demonstrated for large network application as part of this study; and
- Network Reliability Measures. In a network simulation model, (1) route choice must include the reliability measures in a way consistent with mode and other choices, and (2) network path–building algorithms must generate O-D measures, along with average travel time and cost, to feed back to the demand model. Two practical approaches are proposed to estimate variability measures of travel time in the context of network assignment tools. The first exploits trajectory information in micro- and mesosimulation tools; the second employs a robust relation established between the first and second moments of the travel time per unit distance. These methods are fully compatible with the adopted functional form of the highway utility and reliability measures like standard deviation of travel time per unit distance.

The proposed integrated model framework is a demonstration of a trip-based integration of a well-calibrated mode choice model in practice and a simulation-based dynamic traffic micro-assignment model. This framework is sufficiently flexible to incorporate other dimensions (e.g., destination and departure time choices) in addition to the mode choice dimension from the

demand side, and it can be readily extended to an activity-based integration of demand models and an activity-based dynamic traffic microassignment model.

Dynamic mode share and toll road usage results of the proposed integrated model are demonstrated on the large-scale New York metropolitan network. The convergence of the proposed algorithms is also examined. The proposed model uniquely addresses the needs of metropolitan agencies for prediction of mode and path choices and the resulting network flow patterns, and it can evaluate a wide range of road-pricing scenarios on large-scale networks.

Incorporation of Results in Applied Travel Models

Different model structures offer different options for the inclusion of advanced forms of the highway utility function. Although certain components can be incorporated in any properly segmented model, others, such as travel time reliability measures or probabilistically distributed VOT, impose strict constraints on the model structure. The main related issues of incorporation of the proposed form of the utility function are addressed in the following sections.

Transferability of Model Structures and Parameters Between Regions, Choice Contexts, and Studies

The study results have three levels of generalization: (1) understanding of general rules of travel behavior and identification of major impacts and mechanisms leading to conceptual model structures, (2) mathematical structures of associated choice models and associated forms of the highway utility function, and (3) estimated choice models with the obtained values of coefficients and significance of particular variables.

The first two levels of transferability—model approaches and structures—can be effectively generalized. Most of the functional forms for highway utility were statistically significant in such different regions as New York City and Seattle, and there was agreement between major findings based on RP and SP types of data. However, a direct transfer of model coefficient values from region to region, or from choice context to choice context, is not recommended. For different areas, even similar choice contexts such as trip mode choice versus tour mode choice may require a significant rescaling of parameters. It practice, it also may be difficult to ensure exactly the same level of model segmentation and variable definition.

The best way to transfer a model structure from region to region is to reestimate it based on local data using the model specification in the current study as the prototype. In transferability tests (e.g., from New York to Seattle), the absolute majority of model coefficients that were significant for one region were significant for the other region, although the values varied.

A second-best approach is to recalibrate the model on aggregate local data rather than fully reestimating it in a disaggregate fashion. Recalibration can be done after the model has been implemented and the results have been compared to the aggregate targets externally established for each choice dimension. Recalibration and full reestimation differ in that only a subset of parameters (bias constants that do not interact with any person, household, land use, or LOS variables) are allowed to change.

Using Study Results in Applied Forecasting Models

An applied forecasting model must meet certain requirements that in turn impose objective limitations on the functional forms of highway utility, specifically, travel time reliability measures. The research results of this study are grounded in one or more of four applied modeling contexts:

- Aggregate (Four-Step) Demand Models. Although these models offer a limited framework for incorporating congestion and pricing effects, some of the main features of the

highway utility function can be incorporated by including the suggested generalized cost components in the mode choice utilities for highway modes. The mode choice model has to differentiate highway modes by three to four occupancy categories and toll or nontoll route, which would result in six to eight highway modes. After adequate segmentation by trip purpose, income groups, and TOD, several hundred trip tables may be generated. However, any additional segmentation by using person, household, or land use characteristics or adding additional choice models would be impossible.

- ABMs Implemented in a Microsimulation Fashion. These models are characterized by a fully disaggregate structure and rely on individual microsimulation of households and persons. They take full advantage of a detailed level of segmentation by household and person characteristics and can include complicated decision-making chains and behavioral mechanisms. The suggested form of the highway utility can be fully implemented, including route-type, mode, and TOD choices. Variables such as income and parameters like VOT can be continuously distributed to account for unobserved heterogeneity (situational variation).
- Static Traffic Assignment. It is probably impossible to incorporate travel time reliability measures in this framework except by use of simplified proxies. Several simplified approaches can be implemented with these models that are still used by many metropolitan planning organizations and departments of transportation. For example, the perceived highway time concept can be readily incorporated on both the demand and network simulation sides. Some improvements to the current state of the practice can be achieved with a multiclass assignment in which vehicle classes are defined by occupancy, route type, and (possibly) VOT-based groups. This practice, however, may result in more than 20 vehicle classes and long run times for large regional networks.
- DTA with Microsimulation of Individual Vehicles. These models are characterized by a fully disaggregate structure and rely on individual microsimulation of vehicles. Similar to ABMs, they can take full advantage of a detailed level of segmentation by household and person characteristics linked to each vehicle, and they can also incorporate probabilistically distributed VOT to account for unobserved user heterogeneity. With the new technical features described in this study, these models can incorporate the suggested O-D measures of travel time reliability in route choice, as well as generate reliability skims to feed back to the demand model.

The major applications framework for the proposed models primarily takes into account the full regional model framework, although facility- and corridor-level models are also taken into account. For deep understanding and proper modeling of congestion and pricing impacts, a full framework, with chosen and nonchosen alternatives, should be available to both users and nonusers, for which full regional travel data set and model are needed. At both the model estimation stage and the application stage, it is essential to know LOS variables such as travel time, cost, and reliability for nonchoice routes, modes, TOD periods, and destinations.

The most promising long-term direction for DTA modeling is the integration of an activity-based demand model with DTA, in which both models are implemented in a fully disaggregate microsimulation fashion with enhanced typological, temporal, and spatial resolution.

Incorporation of Travel Time Reliability in Operational Models

The incorporation of travel time reliability measures in demand models, and especially in network simulations models, still represents a major challenge, especially if the modeling system is to be practical in terms of run time and data support. In general, there are four possible approaches to quantifying reliability:

- Indirect Measure: This concept is based on statistical evidence that in congestion conditions, travelers perceive each minute with a certain weight. Perceived highway time is not a direct

measure of reliability, but it can serve as a proxy for reliability because the perceived weight of each minute spent in congestion is a consequence of associated unreliability.

- First Direct Measure: Time Variability (Distribution) Measures. This direct approach assumes that several independent measurements of travel time are known that allow for forming the travel time distribution and calculation of derived measures, such as buffer time.
- Second Direct Measure: Schedule Delay Cost. According to this concept, the direct impact of travel time unreliability is measured through cost functions (penalties expressed in monetary terms) of being late (or early) compared with the planned schedule of the activity. This approach assumes that the desired schedule is known for each person and activity undertaken in the course of the modeled period.
- Third Direct Measure: Loss of Activity Participation Utility. This approach assumes that each activity has a certain temporal utility profile and that individuals plan their schedules to achieve maximum total utility over the modeled period, taking into account expected travel times. An extended travel time due to unreliability can be associated with a loss of a participation in the corresponding activity. Similar to the schedule delay concept, this approach suffers from data requirements that are difficult to meet in practice.

Current possibilities for incorporating each approach within the specific frameworks of both demand modeling and network simulation and supporting it with the necessary input data are summarized in Table ES.3.

Summary of Recommended Model Parameters

A summary of the recommended (default) values for all coefficients applied in the highway utility function (Equation ES.2) is given in Table ES.4. These parameters are recommended for use in operational models only if a full disaggregate estimation of regional data cannot be implemented. In that case, careful aggregate validation and calibration of the entire model system, including route-type, mode, and TOD choices, will be needed.

Table ES.3. Incorporation of Travel Time Reliability Approaches in Operational Models

Method	Demand Model	Network Simulation
Perceived highway time	Straightforward and does not require structural changes	Straightforward and does not require structural changes
Time distribution (mean variance)	Straightforward and does not require structural changes	Network route choice has to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D reliability measures need to be generated
Schedule delay cost	Preferred arrival time has to be externally specified for each trip	Network route choice has to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D travel time distributions should be generated either analytically or through multiple simulations
Loss of participation in activities	Temporal utility profiles have to be specified for each activity; entire-day schedule consolidation model has to be applied	Network route choice needs to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D travel time distributions have to be generated either analytically or through multiple simulations

Table ES.4. Recommended Coefficient Values

Travel Purpose	Model Coefficients								Examples of Population and Travel Characteristics			Derived Measures					
	Toll Bias	Time (min)	Distance (mi) Linear	Distance (mi) Squared	Cost (cents)	SD per mi (min/mi)	Exponent for Income	Exponent for Car Occupancy	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Toll Bias Equivalent (min)	Cost Coefficient with Income and Occupancy Effects	VOT ($/h)	VOR ($/h)	Reliability Ratio
To Work and Business	–0.85	–0.0425	0.02024	–0.000266	–1.25	–0.625	0.6	0.8	30,000	1.0	5.0	–0.0465	18.3	–0.0026	10.8	29.1	2.69
									30,000	2.0	5.0	–0.0465	18.3	–0.0015	18.9	50.7	2.69
									30,000	3.0	5.0	–0.0465	18.3	–0.0011	26.1	70.2	2.69
									30,000	1.0	10.0	–0.0500	17.0	–0.0026	11.6	14.6	1.25
									30,000	2.0	10.0	–0.0500	17.0	–0.0015	20.3	25.4	1.25
									30,000	3.0	10.0	–0.0500	17.0	–0.0011	28.0	35.1	1.25
									30,000	1.0	20.0	–0.0552	15.4	–0.0026	12.9	7.3	0.57
									30,000	2.0	20.0	–0.0552	15.4	–0.0015	22.4	12.7	0.57
									30,000	3.0	20.0	–0.0552	15.4	–0.0011	31.0	17.5	0.57
									60,000	1.0	5.0	–0.0465	18.3	–0.0017	16.4	44.2	2.69
									60,000	2.0	5.0	–0.0465	18.3	–0.0010	28.6	76.9	2.69
									60,000	3.0	5.0	–0.0465	18.3	–0.0007	39.6	106.4	2.69
									60,000	1.0	10.0	–0.0500	17.0	–0.0017	17.7	22.1	1.25
									60,000	2.0	10.0	–0.0500	17.0	–0.0010	30.7	38.4	1.25
									60,000	3.0	10.0	–0.0500	17.0	–0.0007	42.5	53.2	1.25
									60,000	1.0	20.0	–0.0552	15.4	–0.0017	19.5	11.0	0.57
									60,000	2.0	20.0	–0.0552	15.4	–0.0010	33.9	19.2	0.57
									60,000	3.0	20.0	–0.0552	15.4	–0.0007	46.9	26.6	0.57
									100,000	1.0	5.0	–0.0465	18.3	–0.0013	22.3	60.0	2.69
									100,000	2.0	5.0	–0.0465	18.3	–0.0007	38.9	104.5	2.69
									100,000	3.0	5.0	–0.0465	18.3	–0.0005	53.8	144.5	2.69
									100,000	1.0	10.0	–0.0500	17.0	–0.0013	24.0	30.0	1.25
									100,000	2.0	10.0	–0.0500	17.0	–0.0007	41.8	52.2	1.25
									100,000	3.0	10.0	–0.0500	17.0	–0.0005	57.8	72.2	1.25
									100,000	1.0	20.0	–0.0552	15.4	–0.0013	26.5	15.0	0.57
									100,000	2.0	20.0	–0.0552	15.4	–0.0007	46.1	26.1	0.57
									100,000	3.0	20.0	–0.0552	15.4	–0.0005	63.8	36.1	0.57

(continued on next page)

Table ES.4. Recommended Coefficient Values (continued)

	Model Coefficients								Examples of Population and Travel Characteristics			Derived Measures					
			Distance (mi)														
Travel Purpose	Toll Bias	Time (min)	Linear	Squared	Cost (cents)	SD per mi (min/mi)	Exponent for Income	Exponent for Car Occupancy	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Toll Bias Equivalent (min)	Cost Coefficient with Income and Occupancy Effects	VOT ($/h)	VOR ($/h)	Reliability Ratio
From Work and Business	–0.95	–0.0425	0.02024	–0.000266	–1.44	–0.545	0.6	0.8	30,000	1.0	5.0	–0.0465	20.4	–0.0030	9.4	22.1	2.34
									30,000	2.0	5.0	–0.0465	20.4	–0.0017	16.4	38.4	2.34
									30,000	3.0	5.0	–0.0465	20.4	–0.0012	22.7	53.1	2.34
									30,000	1.0	10.0	–0.0500	19.0	–0.0030	10.1	11.0	1.09
									30,000	2.0	10.0	–0.0500	19.0	–0.0017	17.6	19.2	1.09
									30,000	3.0	10.0	–0.0500	19.0	–0.0012	24.3	26.6	1.09
									30,000	1.0	20.0	–0.0552	17.2	–0.0030	11.2	5.5	0.49
									30,000	2.0	20.0	–0.0552	17.2	–0.0017	19.4	9.6	0.49
									30,000	3.0	20.0	–0.0552	17.2	–0.0012	26.9	13.3	0.49
									60,000	1.0	5.0	–0.0465	20.4	–0.0020	14.3	33.4	2.34
									60,000	2.0	5.0	–0.0465	20.4	–0.0011	24.8	58.2	2.34
									60,000	3.0	5.0	–0.0465	20.4	–0.0008	34.4	80.5	2.34
									60,000	1.0	10.0	–0.0500	19.0	–0.0020	15.3	16.7	1.09
									60,000	2.0	10.0	–0.0500	19.0	–0.0011	26.7	29.1	1.09
									60,000	3.0	10.0	–0.0500	19.0	–0.0008	36.9	40.3	1.09
									60,000	1.0	20.0	–0.0552	17.2	–0.0020	16.9	8.4	0.49
									60,000	2.0	20.0	–0.0552	17.2	–0.0011	29.5	14.6	0.49
									60,000	3.0	20.0	–0.0552	17.2	–0.0008	40.8	20.1	0.49
									100,000	1.0	5.0	–0.0465	20.4	–0.0014	19.4	45.4	2.34
									100,000	2.0	5.0	–0.0465	20.4	–0.0008	33.7	79.1	2.34
									100,000	3.0	5.0	–0.0465	20.4	–0.0006	46.7	109.4	2.34
									100,000	1.0	10.0	–0.0500	19.0	–0.0014	20.8	22.7	1.09
									100,000	2.0	10.0	–0.0500	19.0	–0.0008	36.3	39.5	1.09
									100,000	3.0	10.0	–0.0500	19.0	–0.0006	50.1	54.7	1.09
									100,000	1.0	20.0	–0.0552	17.2	–0.0014	23.0	11.4	0.49
									100,000	2.0	20.0	–0.0552	17.2	–0.0008	40.0	19.8	0.49
									100,000	3.0	20.0	–0.0552	17.2	–0.0006	55.4	27.3	0.49

(continued on next page)

Table ES.4. Recommended Coefficient Values (continued)

Travel Purpose	Model Coefficients								Examples of Population and Travel Characteristics			Derived Measures					
	Toll Bias	Time (min)	Distance (mi) Linear	Distance (mi) Squared	Cost (cents)	SD per mi (min/mi)	Exponent for Income	Exponent for Car Occupancy	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Toll Bias Equivalent (min)	Cost Coefficient with Income and Occupancy Effects	VOT ($/h)	VOR ($/h)	Reliability Ratio
Nonwork	–1.2	–0.0335	0	0	–0.5228	–0.418	0.5	0.7	30,000	1.0	5.0	–0.0335	35.8	–0.0030	6.7	16.6	2.50
									30,000	2.0	5.0	–0.0335	35.8	–0.0019	10.8	27.0	2.50
									30,000	3.0	5.0	–0.0335	35.8	–0.0014	14.4	35.9	2.50
									30,000	1.0	10.0	–0.0335	35.8	–0.0030	6.7	8.3	1.25
									30,000	2.0	10.0	–0.0335	35.8	–0.0019	10.8	13.5	1.25
									30,000	3.0	10.0	–0.0335	35.8	–0.0014	14.4	17.9	1.25
									30,000	1.0	20.0	–0.0335	35.8	–0.0030	6.7	4.2	0.62
									30,000	2.0	20.0	–0.0335	35.8	–0.0019	10.8	6.7	0.62
									30,000	3.0	20.0	–0.0335	35.8	–0.0014	14.4	9.0	0.62
									60,000	1.0	5.0	–0.0335	35.8	–0.0021	9.4	23.5	2.50
									60,000	2.0	5.0	–0.0335	35.8	–0.0013	15.3	38.2	2.50
									60,000	3.0	5.0	–0.0335	35.8	–0.0010	20.3	50.7	2.50
									60,000	1.0	10.0	–0.0335	35.8	–0.0021	9.4	11.8	1.25
									60,000	2.0	10.0	–0.0335	35.8	–0.0013	15.3	19.1	1.25
									60,000	3.0	10.0	–0.0335	35.8	–0.0010	20.3	25.4	1.25
									60,000	1.0	20.0	–0.0335	35.8	–0.0021	9.4	5.9	0.62
									60,000	2.0	20.0	–0.0335	35.8	–0.0013	15.3	9.5	0.62
									60,000	3.0	20.0	–0.0335	35.8	–0.0010	20.3	12.7	0.62
									100,000	1.0	5.0	–0.0335	35.8	–0.0017	12.2	30.3	2.50
									100,000	2.0	5.0	–0.0335	35.8	–0.0010	19.8	49.3	2.50
									100,000	3.0	5.0	–0.0335	35.8	–0.0008	26.2	65.5	2.50
									100,000	1.0	10.0	–0.0335	35.8	–0.0017	12.2	15.2	1.25
									100,000	2.0	10.0	–0.0335	35.8	–0.0010	19.8	24.6	1.25
									100,000	3.0	10.0	–0.0335	35.8	–0.0008	26.2	32.7	1.25
									100,000	1.0	20.0	–0.0335	35.8	–0.0017	12.2	7.6	0.62
									100,000	2.0	20.0	–0.0335	35.8	–0.0010	19.8	12.3	0.62
									100,000	3.0	20.0	–0.0335	35.8	–0.0008	26.2	16.4	0.62

Note: SD = Standard deviation.

CHAPTER 1

Research Objectives and Main Methodology

Three Levels of Specification

The research agenda for the SHRP 2 C04 project topic required attention to both theoretical and applied perspectives. Succinctly stated, the primary objectives are as follows:

- Theory and Research. Develop mathematical descriptions of the full range of highway users' behavioral responses to congestion, travel time reliability, and pricing; and
- Application in Practice. Provide guidance for incorporating these mathematical specifications into various travel demand models currently in use (and under development), recognizing the complex nature of supply- (network-) side feedbacks (via traffic assignment techniques).

The research was conceptualized in three interconnected levels of behavioral rigor and practical application, with varying levels of sophistication and associated inputs:

- Level 1. Behavior foundations;
- Level 2. Advanced operational modeling (activity or tour based); and
- Level 3. Opportunities for prevailing practice (aggregate trip based).

Because supply–demand interactions are critical for congestion and pricing solutions (including network equilibrium), these interactions offer a second dimension, as reflected in Figure 1.1.

Level 1: Behavior Foundations

The first level, as shown in Figure 1.1, corresponds to behavioral models intended for a deep understanding and quantitative exploration of travel behavior. These models include many kinds of variables, often explicitly controlled under stated-preference (SP) settings (e.g., preferred arrival time and schedule flexibility) and not all of which can be produced by most network or supply-side models (e.g., travel time reliability, particularly in the event of nonrecurring incidents). These models seek to address the full range of possible short- and long-term responses, but they may also focus on select choice dimensions (e.g., route and departure time choices or usual workplace location choice).

Supply-side variables for such models can be based on observed or generated measures (or both) of congestion, reliability, and price (via, for example, an SP survey design). Multiple, repeated observations can be used for direct derivation of reliability measures. Typically, there is no consideration of equilibrium at this stage, and the linkage between the demand and supply sides is essentially one-directional.

Research associated with the widest possible range of behavioral responses is important for the construction of an "ideal" behavioral model. Such a model is free of implementation constraints, but with some simplifying assumptions, it is able to serve as the starting point for operational models. In particular, this exploratory level considers dynamics—adjustments of within-day, as well as day-to-day, time frames: short term (in which case the effect of information could be included), medium term, and long term—as well as the correspondence of the time scale to different choice dimensions. For example, in certain situations for short-term analysis, route choice might be the only relevant dimension, but departure time choice may be equally important for day-to-day and medium- and long-term responses.

Level 2: Advanced Operational Modeling

The second level relates to the emerging set of relatively advanced, yet operational, activity-based models (ABMs) that are integrated with state-of-the-art dynamic traffic assignment (DTA) models for network simulation. These models allow for the incorporation of a wide range of possible short- and long-term responses that are embedded within

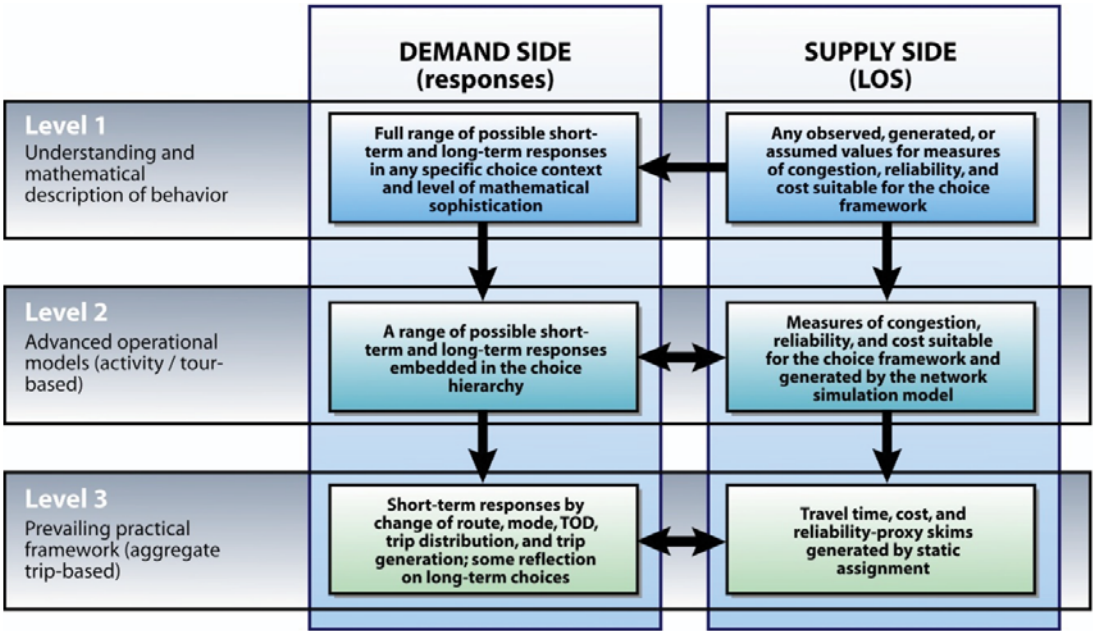

Figure 1.1. Levels of analytic sophistication.

the choice hierarchy. For example, a traveler's acquisition of an E-ZPass or transponder may be linked to his or her subsequent choice of payment type (at the lower level of the behavioral hierarchy). The integrity of operational models requires that each choice dimension should be allocated a proper "slot" in the hierarchy, with upward and downward linkages to related choices. Operational and computing time requirements often limit the total number of choice dimensions and alternatives, but this restriction is lessening with time.

Another relevant constraint in model application is that all measures of congestion, reliability, and price must be compatible with the demand model's specification and can be generated by the network simulation. Moreover, the demand and supply sides should be integrated in an equilibrium setting, which imposes certain limitations on how variables like travel time variability are generated, as direct methods based on multiple observations of the same trip typically are infeasible.

Level 3: Opportunities for Prevailing Practice

The third level relates to the larger number of existing model systems used by most metropolitan planning organizations (MPOs) and state departments of transportation (DOTs) in the form of aggregate trip-based models (frequently referred to as "four-step" models). Although rather restrictive in design, such models are prevalent and offer opportunities for meaningful and immediate contributions to the state of travel demand modeling practice. Although the four-step framework emphasizes short-term responses to highway congestion and pricing policies (including changes in route, mode, and in some cases, departure time, for each trip segment), it allows for the incorporation of some trip distribution and even trip generation effects through generalized cost impedances (mode choice logsums) and accessibility measures (destination choice logsums). The four-step model framework also allows for some indirect reflection of pricing on long-term choices, including workplace location and car ownership. A serious restriction of four-step models, also common to most ABMs in current implementation, is that they rely on static assignment procedures. Static assignments generate only crude average travel time and cost variables, and reliability can be incorporated only through certain proxies.

In this respect, the C04 research project has aimed to push the boundary of the network models to achieve greater behavioral sensitivity within the demand models, along with natural integration of all system components. Although several advanced models and methods exist, they require special data sets and longer run times, along with other use restrictions, many of which are purely technical. For example, DTA at a full regional scale is not yet realistic, although with computational advances and parallel processing opportunities, a dramatic breakthrough may be anticipated in the next 5 to 10 years, or possibly sooner. These constraints on practical applications also relate to the current limitations of demand models in terms of possible number of choice

dimensions and numerical realizations in the microsimulation process.

Incorporation of Results in Applied Travel Models

It is important to note that each level is not independent and disconnected from the others. The team aims to establish a consistent and holistic conceptual framework in which simplified and pragmatic models can be derived from more advanced models, rather than being reinvented (which is probably the current state of the relationship between travel modeling theory and practice). This way, it is believed that the current project could be successful in a very important respect: by bridging the gaps between theory and practice to the greatest extent possible.

The major framework for the discussion of the proposed models primarily considers the full regional model framework, although facility- and corridor-level models are also considered. This focus also has a consequence for the analysis of the existing data sets and their possible use in the current research. It is based on the recognition that for a deep understanding and proper modeling of congestion and pricing impacts, a full framework, with chosen and nonchosen alternatives, should be available to both users and nonusers, for which a full regional travel data set and model is needed. It is essential to know, both at the model estimation stage and the application stage, level-of-service (LOS) variables such as travel time, cost, and reliability for nonchoices routes, modes, time-of-day (TOD) periods, and destinations. This holistic framework is generally missing in simplified models and surveys, which limits their utility for this research.

According to the adopted levels of sophistication, the research results of this C04 topic are grounded in one or more of three applied modeling contexts:

- Aggregate (Four-Step) Models and Static Assignment Tools. In general, these models offer a limited framework for the incorporation of congestion and pricing effects. In particular, it is problematic to incorporate travel time reliability measures in these frameworks. However, the team has formulated several simplified approaches that can be implemented within these models as they are still in use by many MPOs and DOTs. For example, the perceived highway time concept can be readily incorporated on both the demand and network simulation sides.
- SP-Based Models. Most advanced behavioral models are primarily intended to provide insights into individual travel behavior. These models (especially in an SP setting) can include additional behavioral variables related to the dynamics of decision making (e.g., previously used route or option), referencing mechanism (e.g., travel time savings versus the actual trip time), flexibility of the work schedules, and preferred arrival time. Usually, however, these network or supply-side variables are not easily simulated in model application; thus, the corresponding model cannot be directly used in the framework of an applied regional travel model.
- ABMs. These advanced applied travel models and network simulation tools are characterized by a fully disaggregate structure and rely on individual microsimulation. They take full advantage of a detailed level of segmentation by household and personal characteristics and can include complicated decision-making chains and behavioral mechanisms. They are, however, limited to only forecastable input variables.

State of the Art and Practice in Modeling Congestion and Pricing

Addressing Impacts of Congestion

Many of the modeling aspects of concern for this research are generally associated with travel behavior and would be relevant for any travel model improvement. In any travel model, there are certain time–cost trade-offs that are not necessarily related to highway pricing and congestion. And practically, any travel model would benefit from a fuller set of behavioral responses and associated choices. The C04 research, however, is specially designed to substantially extend our understanding and ability to model the impacts of transportation pricing and congestion. For this reason and due to this focus, the research was not intended to result in a general travel model improvement guide, although all recommendations and developed approaches were arrayed as to their relevance for the general state of the art and the practice of the profession.

The most interesting and unique aspect of the research is the focus on impacts of congestion and associated pricing on traveler response and transportation system performance. Almost all existing models are already sensitive to congestion through (average) travel time variables, at least in their assignment and mode choice components. So, what is special in congestion that requires a special consideration beyond the framework of conventional model structures and approaches?

Travel time reliability is one of most important aspects investigated in this research that has been generally recognized as a critical missing component in the previous generation of

travel demand models. Congestion is associated not only with longer average travel times, but also with higher levels of unreliability (unpredictability) of travel times, which is what makes it so onerous to highway users. A great deal of the project effort was related to the measurement and incorporation of reliability in model structures. But important as it is, reliability is not the only additional issue or variable that needs to be added to the existing travel models in order to have a better accounting for congestion.

The team believes that a deeper understanding of the effects of congestion on travel behavior should include several additional considerations:

- Perceived Highway Travel Time. The practice of using differential weights for different travel time components has been universally accepted for transit modeling. Transit in-vehicle time, walk time, and wait time are perceived by the riders differently. The corresponding estimated utility coefficients normally range between 1.0 and 4.0, with the highest weights associated with waiting time under uncertain conditions. But there has been no parallel effort to estimate perceived highway time as a function of highway LOS, which has always been assumed to be a totally generic variable in both route choice and mode choice contexts, as well as in any subsequent use of mode choice logsums or generalized cost in trip distribution and upper-level models. However, a behavioral analogue between an uncertain waiting time for an unreliable transit service and for being stuck in a car in a traffic jam is appealing. The team believes that the idea of a perceived time structure (e.g., by travel speed categories) might be beneficial from both a theoretical and a practical modeling perspective. The introduction of weighted highway time would improve mode choice models and allow for the elimination of some mysterious constants (such as "rail bias for trips to the central business district").
- Different Pattern of Highway User Behavior in Presence of Unpredictable Travel Times. Another assumption underlying conventional modeling approaches that becomes unrealistic under congestion conditions is that travelers (specifically highway users) possess full information about all possible routes and modes and make rational decisions. In behavioral terms, congestion and the associated unpredictability of travel times lead to travelers making many irrational decisions based on intuition and past experience that might not be relevant for the current situation. In practical modeling terms, it might be expected that the associated choice models would have relatively smaller coefficients for travel time and cost (more random behavior and regardless of value of time [VOT]) compared with models estimated for uncongested areas where travel time is predictable. As a result, in a route choice framework large deviations from the shortest path might be expected. This general pattern might be affected by the travel information system, and more so as congestion creates demand for real-time information. The impact of intelligent transportation systems in this context represents an interesting research challenge (though not in the current project's scope).
- Disequilibrium (Lagged Equilibrium) Between Travel Demand and Supply. Another interesting and less-investigated consideration relates to the equilibrium formulation. It is generally recognized that travel models should reach a perfect (simultaneous) equilibrium between the demand and supply sides. A corresponding theory and effective algorithms have been well established for aggregate four-step models. Although it is more empirical with microsimulation ABMs, the intention, however, is still to reach a perfect equilibrium. It is interesting to note that integrated land use and transportation models have never used a concept of static equilibrium because the land use and transportation responses belong to different time scales. Most integrated land use and transportation models incorporate the concept of lagged equilibrium. In reality, there are numerous and very different time scales within a travel demand model itself. In the presence of congestion that makes travel time unstable, the process of traveler learning and adaptation associated with reaching equilibrium becomes longer and fuzzier. Research has shown that it might be beneficial to revise the formulation of transportation system equilibrium accordingly.

Modeling Toll Roads and Managed Lanes

In the same way that the experience of congestion is not only about longer travel times, priced highway facilities themselves are more than just roads with better travel times available at additional cost. Important qualities of a tolled facility, as well as the traveler's perception of them, include many other aspects that make priced highway facilities qualitatively different from free highways, to such an extent that they can be better modeled as a different mode, rather than merely a different route in the network (Spear 2005; Erhardt et al. 2003). Since the difference between priced and free highway facilities from the traveler's perspective is probably not as great a difference as between highway and transit modes, in the mode choice structure, the choice between toll and nontoll routes (preroute choice) is normally placed in the lowest level of nested structure. However, it is important to keep the toll–nontoll choice as a discrete choice in order to allow for the inclusion of various utility components and biases. The assignment framework is more limited in this respect.

The following specific preroute factors (beyond travel time and cost) should be considered and estimated for all types of priced facilities:

- Reliability. Toll roads and managed lanes, especially those with variable and real-time pricing, are perceived as reliable modes of transportation; congested free roads are inherently unpredictable (Bates et al. 2001; Brownstone and Small 2005; Small et al. 2005);
- Safety. Some drivers may perceive priced facilities like mainline toll (high-occupancy toll [HOT]) lanes to be safer than the mainline because of the separation from the other lanes of travel, and specifically from trucks (Brownstone et al. 2003); and
- Carpooling Opportunity. Several additional factors come into play if pricing or traffic restrictions are differentiated by car occupancy. Different forms of high-occupancy vehicle (HOV)–HOT lanes have been recently applied in many states. Valuable experience with these lanes has already been accumulated, and a significant body of model estimation work and research has been published. The team plans to cooperate closely with the ongoing NCHRP Project 8-36B, Task 52, Changes in Travel Behavior/Demand Associated with Managed Lanes.

Highway Utility Forms in Different Demand Choice Frameworks

Highway Utility Components

Highway travel utility is the basic expression of combining various LOS attributes and costs as perceived by the highway user. It is directly used in the highway trip route choice; for example, it is used between the managed lanes and general-purpose lanes on the same facility. It also constitutes an essential component in mode and TOD choice utilities. The form of highway utility function is also important for modeling other (upper-level) travel choices because it serves as the basis for accessibility measures. Thus, it is essential to explore the formulation of highway travel utility and its components first, having in mind a simplified framework of route choice in the highway network, because the complexity builds when additional choice dimensions are considered.

In most travel demand models, including those developed for practical and research purposes, the highway utility (U) takes the following simple form:

$$U = a \times T + b \times C \tag{1.1}$$

where

T = travel time;
C = travel cost;
$a < 0$ = coefficient for travel time;
$b < 0$ = coefficient for travel cost; and
a/b = VOT.

Coefficients for travel time and cost normally take negative values, reflecting the fact that travel in itself is an onerous function necessary only for visiting the activity location. Thus, the travel utility is frequently referred to as the "disutility" of travel. Some research has questioned the negative character of travel utility in some contexts. In particular, a positive travel utility was associated with long recreational trips on weekends (Stefan et al. 2007). Also, an interesting effect was observed for commuting trips for which commuters seem to prefer a certain minimum time and are not interested in reducing it below this threshold (Redmond and Mokhtarian 2001).

The standard representation of highway travel utility as a linear function of two variables with constant coefficients is extremely simplified. A great deal of the present research effort has been devoted to the elaboration of this basic form in the following ways:

- Investigation of the highway user perception of travel time by congestion levels. This means that a simple generic coefficient for travel time could be replaced with coefficients differentiated by congestion levels;
- Inclusion and estimation of additional components, of which travel time reliability has been identified as the most important one. With respect to average travel time and cost, reliability is seen as an additional and nonduplicating term; and
- Testing more complicated functional forms that are nonlinear in time and cost, as well as account for randomly distributed coefficients or VOT (in addition to any explicit segmentation accounting for the observed user heterogeneity). With these enhancements, VOT is not assumed as a constant, but becomes a varying parameter depending on the absolute values of time and cost, as well as reliability.

As a working model, the team has adopted the following general expression for the highway travel utility, and explored it component-by-component over the course of the research:

$$U = \sum_{k=1}^{5}[a_k \times \varphi_k(T_k)] + \sum_{m=1}^{3}[b_m \times \phi_m(C_m)] + \sum_{n=1}^{3} c_n R_n \tag{1.2}$$

where

$k = 1$ represents the uncongested highway travel time component;
$k = 2$ represents the congested highway travel time component (extra delay);
$k = 3$ represents parking search time;

$k = 4$ represents walk access or egress time (e.g., from the parking lot to the trip destination);
$k = 5$ represents extra time associated with carpooling (i.e., picking up and dropping off passengers);
$T_k =$ (average) travel time by component;
$m = 1$ represents highway toll value;
$m = 2$ represents parking cost;
$m = 3$ represents vehicle maintenance and operating cost;
$C_m =$ travel cost value by component;
$n = 1$ represents disutility of time variation (first measure of reliability);
$n = 2$ represents schedule delay cost (second measure of reliability);
$n = 3$ represents utility of (lost) activity participation (third measure of reliability);
$R_n =$ reliability measures by component;
$a_k, b_m, c_n =$ coefficients to be estimated; and
$\varphi_k(\ldots), \phi_m(\ldots) =$ functions for nonlinear transformation of time and cost variables.

This formulation makes it more difficult to calculate VOT, although the calculation is still possible, in the same way that value of reliability (VOR) can be calculated for the first type of reliability measure (assuming that this reliability measure is in minutes). VOR essentially represents travelers' willingness to pay for reduction in travel time variability in the same way as VOT represents their willingness to pay for (average) travel time savings. More specifically, VOT (in the context of willingness to pay tolls for saving time in congestion conditions) can be calculated by the following general formula:

$$VOT(T_2, C_1) = \frac{\partial U/\partial T_2}{\partial U/\partial C_1} = \frac{a_2 \varphi_2'(T_2)}{b_1 \phi_1'(C_1)} \tag{1.3}$$

A similar calculation can be implemented for VOR. With nonlinear transformation functions, VOT and VOR are no longer constant values. They now depend on the absolute values of time and cost variables at which the derivatives of the transformation functions are taken.

The innovative components that relate to perceived highway time, travel time reliability, and nonlinear transformations are discussed in the subsequent sections. It should be noted that some components, specifically perceived travel time and the three reliability measures, might be correlated statistically (and also conceptually duplicative to some extent). Thus, it is highly improbable that the entire formula would ever be applied. Instead, it serves as a conceptual framework for which particular structures can be derived and tested statistically against each other.

Perceived Highway Time

Perceived transit time has been recognized and routinely used in travel models for some time. For example, in most mode choice models and transit assignment algorithms, out-of-vehicle transit time components like *wait* and *walk* are weighted compared with in-vehicle travel time. It is not unusual to apply weights in the range of 2.0 to 4.0, reflecting the fact that travelers' perception of out-of-vehicle time is perceived as more onerous than in-vehicle time.

Contrary to transit modeling practice, practically all travel models include only a generic highway time term; that is, the same coefficient is applied for each minute of highway time regardless of the travel conditions. However, there is some compelling statistical evidence that highway users perceive travel time differently by congestion levels. For example, it is intuitive and behaviorally appealing that highway users driving in congested conditions might perceive the longer travel time as an additional delay or penalty on top of the anticipated free-flow (or some expected reasonable) time. Thus, the research has explored a segmentation of travel time coefficients by congestion levels, expecting that the time spent in congestion conditions has a larger disutility. A larger disutility associated with congestion would have at least two behavioral interpretations:

- Negative psychological perception (similar to the weight for walking to or waiting for transit service); and
- Simplified operational proxy for reliability (that should be explored in combination with the explicit reliability measures).

Several related research works report statistical evidence of high perceptional weights that highway users put on travel time in congested conditions (Small et al. 1999; Axhausen et al. 2007; Levinson et al. 2004; McCormick Rankin Corp. and Parsons Brinckerhoff 2008). Multiple indications in recent analyses of travel surveys suggest that the perception of the time saved by respondents in revealed preference (RP) surveys is about double the actual measured time saved (Small et al. 2005; Sullivan 2000). This might well be a manifestation that in the RP framework travelers operate with perceived travel times, in which time spent traveling through congested segments is psychologically doubled.

Major Focus for Improvement for Demand Analysis

The C04 research into opportunities for extended research and improvements in travel demand analysis with respect to pricing and congestion has focused on improving the

following key structural dimensions or components of demand models:

- Primary Choice Dimensions. Models are grouped by primary choice dimensions that relate to congestion and pricing (e.g., route, TOD, and mode choices). It is shown how improved specifications of models can be effectively used in models that relate to the upper-level choices in the individual travel decision-making hierarchy, including destination choice, tour and trip generation, and household car ownership. Joint choice formulations for route and mode, route and TOD, and mode and TOD, as well as route mode and TOD, have also been investigated;
- Specification of Highway Utilities. For each choice model, the basic specification that includes average time and cost have been investigated, first including possible linear and nonlinear specifications. The main points of improvements include possible nonlinear effects and trip length scaling;
- Segmentation Options. For each model various segmentation strategies are tested, including a full segmentation of the choice model by travel purpose and partial segmentation of travel time by congestion levels or travel cost (by income group and occupancy), or both. Main points of improvement studied include substantiation of the concept of perceived highway time with significantly different VOT by congestion levels as a proxy for travel time reliability;
- Income Effects. Special focus is on the impact of household income and corresponding functional specifications of the highway utility. This includes segmentation of some coefficients by income group, using income-specific constants and scaling of the cost by income (as a continuous variable). Alternative approaches are compared in a systematic way and recommendations for best functional forms are made;
- Car Occupancy. Special focus is on the car occupancy effects. Travel forecasting models commonly assume that travel costs should be divided by vehicle occupancy, with the implicit assumption that those costs are shared among those traveling together. This hypothesis is questioned, and alternative formulations are explored, including segmentation of cost or time coefficients (or both) by occupancy, occupancy-specific bias constants, nonlinear scaling occupancy effects, and spate analysis of interhousehold and intrahousehold carpools. Alternative approaches are compared in a systematic way, and recommendations for best functional forms are made; and
- Incorporation of Travel Time Reliability. Significant focused effort has been made to test different travel time reliability measures and incorporate them in the route, mode, and TOD choice utility expressions. The results represent cutting edge research and provide valuable insights into travelers' decision-making process and preferences. The estimated models provide VOR estimates that along with the VOT estimates portray travelers' willingness to pay for different types of highway improvements.

Major Focus for Improvement for Network Simulations

The C04 research project has addressed recent advances in traffic microsimulation tools, dynamic equilibrium algorithms and implementation techniques for large-scale network applications, richer behavioral representation in network models, and ways to generate travel time distributions and reliability measures. Salient points of the research include the following:

- Need for Microsimulation. Capturing user responses to pricing and reliability is best accomplished through microsimulation of individual traveler decisions in a network platform. These responses must be considered in a network setting, not at the facility level, and the time dimension is essential to evaluating the impact of congestion pricing and related measures; hence, a time-dependent analysis tool is required. Microsimulation of individual traveler choices provides the most general and scalable approach to evaluate the measures of interest in this study;
- More Robust DTA Required. Simulation-based DTA models have gained considerable acceptance in the past few years, yet adoption in practice remains in its infancy. The current generation of available models only considers fixed, albeit time-varying, origin–destination trip patterns. Greater use and utility will result from consideration of a more complete set of travel choice dimensions and incorporation of user attributes, including systematic and random heterogeneity of user preferences;
- Improved Algorithms for Regional Scale Modeling. Such algorithms for finding equilibrium time-varying flows have been based on the relatively inefficient method of successive averages, and its implementation in a flow-based procedure did not scale particularly well for application to large metropolitan networks. New implementations of the method of successive averages and other algorithms that exploit the vehicle-based approach of simulation-based DTA have been proposed and demonstrated on large actual networks;
- Traveler Heterogeneity. Incorporating heterogeneity of user preferences is an essential requirement for modeling

user responses to pricing in a network setting. New algorithms that exploit nonparametric, multicriteria shortest-path procedures allow VOT (which determines users' choice of path and mode in response to prices) to be continuously distributed across users. Efficient implementations of these algorithms have been demonstrated for large network application as part of this study; and

- Network Reliability Measures. Most simulation models do not produce reliability estimates of travel time along network links and paths. Two practical approaches were formulated and explored as part of this work to estimate variability measures of travel time in the context of network assignment tools. The first exploits trajectory information in micro- and mesosimulation tools; the second employs a robust relation established between the first and second moments of the travel time per unit distance. These are illustrated for application in conjunction with network evaluation tools.

CHAPTER 2

Review and Selection of Data Sources

This chapter describes the main data sources selected and used for the estimation of models that are described in detail in Chapter 3. To support the specification and estimation of advanced modeling components in C04, the team required certain types of information that are typically not available, but that were needed to overcome certain critical data issues surrounding the joint analysis of congestion and pricing. These critical data issues included

- Lack of observed data on travel time variability and reliability;
- Correlation between key time and cost variables;
- Bidirectional causality between travel times and demand;
- Lack of actual pricing options; and
- Lack of validation for stated preference (SP)–based methods.

Although a wide range of potential data sets were identified and evaluated in Phase 1, many of them shared the limitations listed above. As a result, the decision was made to focus the model estimation work to be done in C04 on the relatively few and most robust of the identified data sources. These are logically grouped by their main characteristics:

- Revealed preference (RP) data on travel demand;
- Network level of service (LOS) and reliability measures;
- SP survey data; and
- Experimental travel data.

Revealed Preference Data on Travel Demand

New York Household Survey

The Regional Travel–Household Interview Survey (RT-HIS) that was used to develop the New York Metropolitan Transportation Council's best practice model (NYBPM) was conducted over a 1-year period in 1997 and 1998. Almost 11,000 households were included in the sample obtained from the 28-county tristate (New York, New Jersey, and Connecticut) NYBPM modeling area. A 1-weekday day travel–activity diary was obtained for each household member. The survey data were recently reweighted and expanded for a base year 2005 update of the NYBPM. While it includes data on amount of tolls paid, it does not include choice of specific tolled facilities or routes. The RT-HIS data, combined with NYBPM network-generated travel time measures, were used directly in the C04 research to estimate congestion and pricing impacts with respect to daily activity patterns, mode, occupancy, destination, and time of day (TOD).

New York Surveys of Toll Facility Users

There are two important recent sets of large-sample origin–destination (O-D) survey data for all of the tolled facilities operated by the New York Metropolitan Transit Authority (MTA) (2006) and by the Port Authority of New York and New Jersey (PANYNJ) (2007). Both include weekday and weekend travel by auto drivers and provide observed characteristics of tolled-crossing users and their trips by E-ZPass, cash, and discounted classes. Both sets of surveys provide the basic segmented traveler, household, and trip data that are critical to modeling. These survey data could support the modeling of transponder acquisition, including the use of data obtained from cash users regarding why they did not use E-ZPass. For the MTA and PANYNJ tolled crossings, detailed traffic count data by toll class and 15-minute intervals are available to support the O-D analysis for these facilities.

Seattle Household Activity Survey

The Seattle Household Activity Survey was carried out for the Puget Sound Regional Council (PSRC) and Washington State Department of Transportation in 2006 (Cambridge Systematics, Inc., Mark Bradley Research and Consulting

et al. 2007). The survey was based on a 2-day travel and activity diary and was carried out on nearly 4,000 households. The households were selected on a geographically stratified basis to enrich the sample in regions with high transit accessibility and opportunities for nonmotorized travel. Intercept samples for ferry riders and park-and-ride users were also included. This survey data set has been used as the basis for activity-based model development at PSRC.

For the C04 project, the Seattle RP data were analyzed at both the tour and trip levels to create models of TOD choice and mode choice. It is important to note that there were no tolled facilities (apart from one bridge) in the Puget Sound region at the time of the survey, so the data is not informative for RP analysis of congestion pricing. The tour file has records for 26,950 tours, with data for over 300 variables, including data on highway travel times in both tour directions for 17 periods during the day. The trip file has records for 73,963 trips, with the same highway LOS variables that are included for tours (but only for the trip direction).

Generation of Network Level of Service and Reliability Measures

Standard Skimming Procedures and Network Level of Service Variables

The standard components of the highway utilities terms can be derived from the standard network skimming procedures found in commercial travel demand forecasting software that generates O-D matrices of travel time LOS and costs. Based on the results of a static and general or user equilibrium assignment procedure, for each O-D pair these methods generate a single value for each of the standard set of fixed highway time and cost measures, which are interpreted as expected or average values, and typically include the following:

- Total travel time;
- Vehicle operating cost (or distance-based function); and
- Toll or other road user costs.

As discussed in the context of the RP model estimation work reported in Chapter 3, some additional measures can be generated with these standard methods that can further specify highway utilities with respect to evaluation of congestion and pricing conditions, including travel times segmented by LOS, speeds, or roadway type (or some combination of these factors).

Like the more conventional measures, these augmented measures remain a single expected fixed value; they do not directly capture any measure of the variability of highway travel conditions that is seen to affect travel choices and are associated with reliability in particular.

Method for Generating Travel Time Distributions

The fact that a distribution of travel times is not generated by standard highway assignment software has two important implications for the research done in the C04 project:

- First, it means that measures of network travel time reliability taken from such distributions are not available to be used in conjunction with RP survey data to estimate models with measures of travel time reliability incorporated in the utilities for highway travel choices. As a result, a vast majority of the models with travel time reliability measures have been estimated in SP settings, in which travel time distributions are predefined as part of the hypothetical choice set respondents consider; and
- Second, without the ability to simulate travel time reliability from standard network assignment procedures, it is not possible to generate these inputs as part of the application of models that include estimated sensitivities to travel time reliability for policy or project forecasting.

Consequently, for the C04 project, a special set of methods was developed for synthesizing a distribution of consistent path-dependent O-D travel times from the distributions of modeled link traffic volumes. This method allows for creating so-called reliability skims that have been used with both the New York and the Seattle-area RP survey data for the estimation of models with a travel time reliability component.

Documentation of the data and methods developed to create these LOS distribution skims needed for the analysis of reliability with RP survey data is provided in Appendix A.

Stated Preference Data

Stated Preference Extension of Seattle Household Survey

For the 2006 PSRC Seattle Household Activity Survey described in this chapter, respondents who had reported trips in relevant transit and highway corridors were selected to participate in a follow-on SP survey. Customized SP scenarios were created based on the reported trip and mailed to respondents. The SP survey was designed by Cambridge Systematics and Mark Bradley Research and Consulting. There were two SP experiments: one was tied to mode choice (bus, rail, and auto), and the other related to TOD tolling on major highways. For the C04 study, the focus was on the latter tolling experiment.

CHOICE EXERCISE 5

Suppose these were your transportation options for your trip from:
a place for conducting personal business at 24TH NE in AUBURN to a workplace at 3401 AUBURN WAY N in AUBURN

	OPTION A	OPTION B	OPTION C	OPTION D
Travel Route	You Travel During the Off-Peak Period (9:00 a.m. to 3:00 p.m. or 7:00 p.m. to 6:00 am)		You Travel During the Peak Period (6:00 a.m. to 9:00 a.m. or 3:00 p.m. to 7:00 pm)	
	Free Route	Toll Route	Free Route	Toll Route
Distance Traveled on Your Trip	8 miles	6 miles	9 miles	6 miles
Usual Travel Time of Your Trip	17 minutes	14 minutes	17 minutes	16 minutes
Reliability: You will be more than 15 minutes late...	1 time out of every 20 times you travel	2 times out of every 20 times you travel	3 times out of every 20 times you travel	1 time out of every 20 times you travel
Toll Cost of Your Trip	$0.00	$2.00	$0.00	$0.75
Which of the four options above would you choose? (Please circle one)	OPTION A	OPTION B	OPTION C	OPTION D

Figure 2.1. Sample SP scenario.

A sample SP scenario is shown in Figure 2.1. Each scenario included four choice alternatives:

- Travel on a free route outside peak periods;
- Travel on a tolled route outside peak periods;
- Travel on a free route during peak periods; and
- Travel on a tolled route during peak periods.

Thus, these data allow estimation of a joint model of route type (tolled versus nontolled) and departure time (peak versus off peak).

In addition to the toll and travel time variables, which are included in all SP experiments of this type, this experiment had two additional variables of interest:

- Distance Traveled. Because the free route may be an entirely different road than the tolled route, there may be a significant difference in terms of distance. In typical RP data, distance is so highly correlated with travel time that it is not feasible to estimate separate time and distance coefficients. This SP data allow the team to estimate such an effect; and
- Reliability of Travel Time. Here, a significant extra delay was defined as "more than 15 minutes late" (beyond the usual travel time), and the scenarios were varied in terms of how often such a delay occurs, allowing the team to estimate the effect of the frequency of delay.

San Francisco Cordon Pricing Stated Preference Survey

The San Francisco County Transportation Authority has recently considered the possibility of implementing cordon pricing around specific areas of downtown San Francisco, California. With Federal Highway Administration (FHWA) funding, an SP survey was carried out in 2007 to aid in modeling the effects of such a policy and set effective levels of cordon charge to influence traffic levels at different times of the day. Auto travelers to downtown were intercepted and participated in a web-based SP interview. The experiment was designed by Mark Bradley and Resource Systems Group (RSG), and the survey was carried out by RSG. An example choice screen from the survey is shown in Figure 2.2.

Each scenario includes four choice alternatives:

- Travel by auto and pay the cordon charge before the peak period;
- Travel by auto and pay the cordon charge during the peak period;
- Travel by auto and pay the cordon charge after the peak period; and
- Travel by public transit.

In contrast to the previous SP example from Seattle, this experiment did not include a nontolled auto alternative,

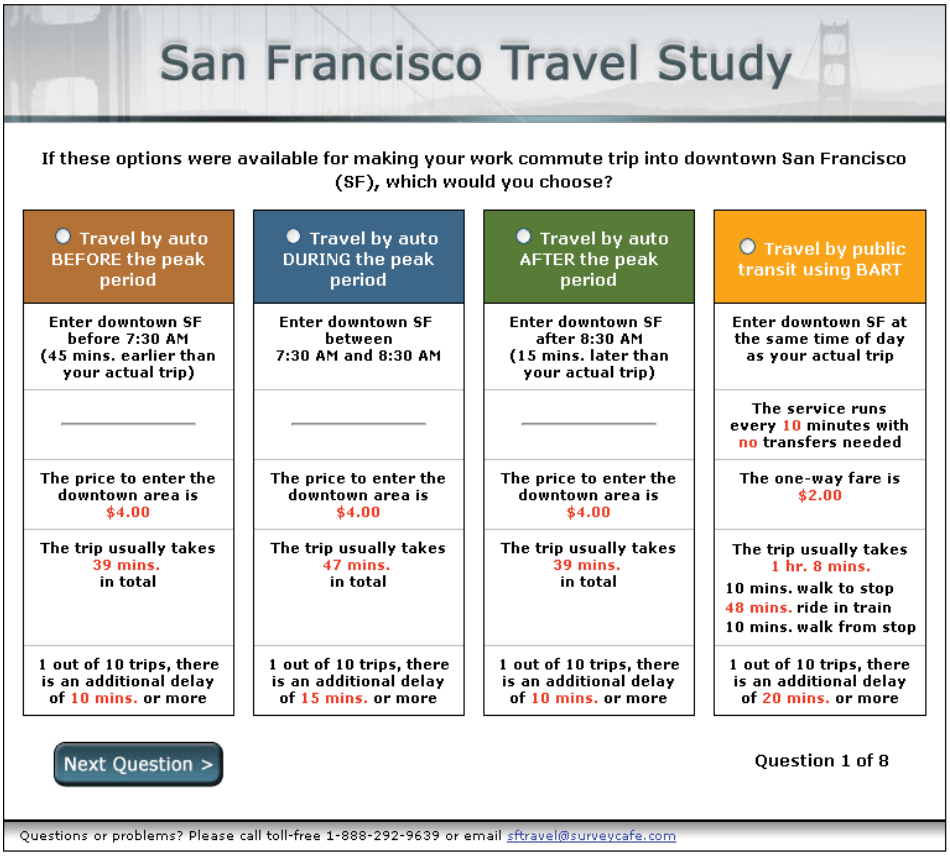

Figure 2.2. Sample choice screen from San Francisco SP survey.

because in the context of cordon pricing, that would mean not traveling to downtown San Francisco at all. (Additional survey questions about that possibility were asked, but they were not analyzed as part of the C04 project.) However, a transit option was included, both because transit to downtown San Francisco is a viable alternative and because part of the stated reason for cordon pricing would be to provide funding to maintain and improve transit services. Thus, the data from this study are suitable for estimating joint models of departure time choice and mode choice.

For the auto alternatives, the variables used for this study were similar to those used for the Seattle SP described in the previous section, except that

- The definition of the peak period used for a given respondent was customized based on their actual departure time, and the duration and timing of the peak pricing period was varied across respondents, allowing a more detailed analysis of departure-time shifting behavior;
- For a given respondent, the effect of reliability was measured by fixing the frequency of delay and varying the length of the delay across the alternatives. This is the opposite of how it was presented in the Seattle SP survey. Frequency was varied randomly across respondents, with "1 out of 10 trips" used for half of the sample and "1 out of 5 trips" used for the other half; and
- All three auto alternatives involved using the same route, so there was no difference in distance.

Los Angeles County Managed-Lane Stated Preference Survey

The County of Los Angeles, California, is considering introducing new managed lanes (express and high-occupancy toll [HOT] lanes) in specific freeway corridors. As part of the preparatory research, an SP experiment was carried out in 2009. Residents of relevant areas were contacted by telephone and recruited if they had made a recent trip by auto using one of the relevant freeways. They were asked for key details of their trip, mailed an SP questionnaire with customized choice scenarios, and then contacted again by telephone to retrieve the responses. The SP experiment was designed by Mark Bradley and PB Americas, and the survey was conducted by Corey, Canapary and Galanis.

An example choice scenario is shown in Figure 2.3. If a person indicated he or she would travel in the off-peak period, a

Option A Express Lane Peak Period	Option B Express Lane Off-Peak Periods	Option C Free Lanes Peak Period	Option D Free Lanes Off-Peak Periods	Option E New Bus Service via Express Lane
Time of Day: *Between 6:30 am and 9:30 am*	Time of Day: *Either before 6:30 am or after 9:30 am*	Time of Day: *Between 6:30 am and 9:30 am*	Time of Day: *Either before 6:30 am or after 9:30 am*	Time of Day: *Any time of the day*
Fuel Cost & Parking Cost: *Same as now*	Fuel Cost & Parking Cost: *Same as now*	Fuel Cost & Parking Cost: *Same as now*	Fuel Cost & Parking Cost: *Same as now*	Transit Service Runs: ***Every 30 minutes***
Toll Charge: ***$0.50***	Toll Charge: ***$0.40***	Toll Charge: *No toll charge*	Toll Charge: *No toll charge*	One-way Fare: ***$5.00***
Travel Time: ***10 minutes***	Travel Time: ***10 minutes***	Travel Time: ***30 minutes***	Travel Time: ***18 minutes***	Travel Time: ***15 minutes*** *door to door, including* ***2 transfer(s)***

If these options had been available for your actual trip, which ones would you have been most likely to use? Please **write in a 1 next to the most attractive option** for your trip, and then **write in a 2 next to the second most attractive option** for your trip.

___ **Option A** (Express Lane during the Peak period)
___ **Option B** (Express Lane, during Off-Peak periods)
___ **Option C** (Free lanes during the Peak period)
___ **Option D** (Free lanes during Off-Peak periods)
___ **Option E** (By transit)

Figure 2.3. Sample choice scenario from the Los Angeles County SP survey.

follow-on question was included to ask if the person would travel before or after the peak. So, there were effectively seven choice alternatives:

- Use the express lane during the peak period;
- Use the express lane before the peak period;
- Use the express lane after the peak period;
- Use the existing free lanes during the peak period;
- Use the existing free lanes before the peak period;
- Use the existing free lanes after the peak period; and
- Use a new bus service via the express lane (in any period).

With these alternatives, it is possible to estimate a joint model along three separate dimensions—route-type choice (tolled versus nontolled), departure time choice, and mode choice—thus combining the scope of the two preceding examples. As in the San Francisco SP example, the definition of the peak pricing period was varied systematically across respondents, and customized somewhat to be relevant for each respondent's actual time of travel, allowing detailed analysis of departure-time shifting. In contrast to the two preceding examples, no explicit travel time reliability variable was included in the scenarios. This decision was made intentionally, because the forecasting framework in which the models will be applied does not include measures of reliability, and including such a variable could influence the estimate for the main travel time coefficient.

Experimental Data

Seattle Traffic Choices Study

The Seattle region currently does not have tolled or priced facilities that would provide much useful data for RP modeling of congestion pricing effects. It does, however, have one unique data set from a recent experiment (Puget Sound Regional Council 2008) that served as one of the principal data sets used for the C04 model estimation. In this experiment, recruited households were given a real monetary budget, and money was deducted from the account every time they used certain roads at certain times of the day and week. Respondents were given a pricing schedule and map, as well as an in-vehicle meter that showed the price whenever they were being charged (Puget Sound Regional Council 2008).

Almost 300 households participated in the Traffic Choices Study for a period of more than 1 year. During that period, GPS data were collected for all trips made in the respondents' vehicles, covering the time span before, during, and after a period when experimental distance-based pricing was administered (for those respondents only). At the beginning of the study each household was given a monetary budget and a schematic pricing chart, with the per mile charge varying by facility type, day of week, and TOD (see the pricing chart in Figure 2.4). Every time one of the household's vehicles

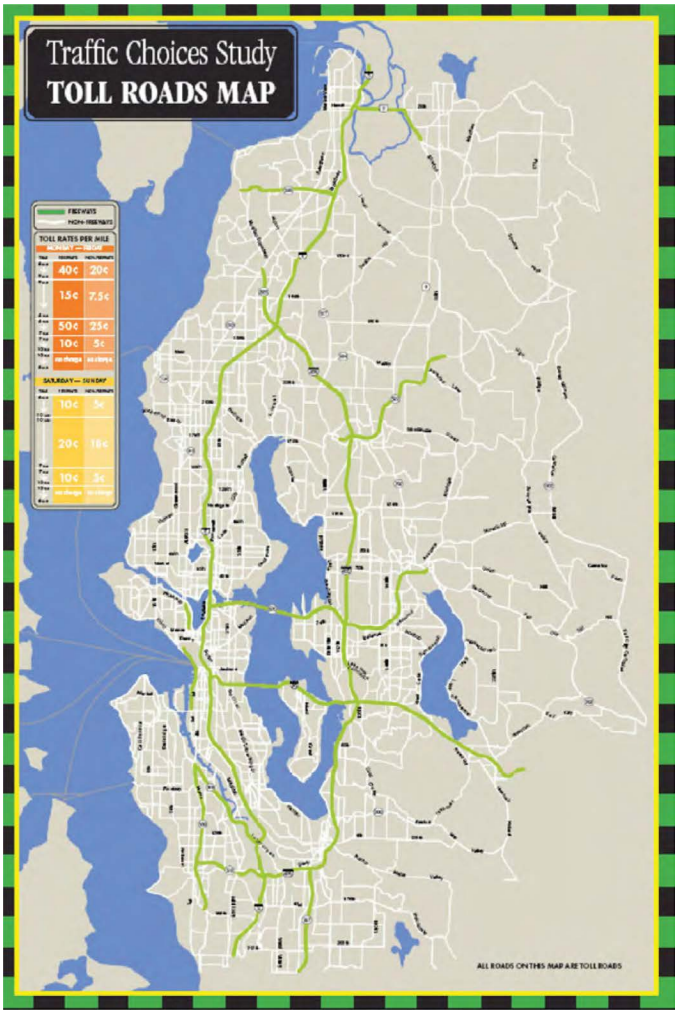

Figure 2.4. Schematic pricing chart in the Seattle Traffic Choices Study.

drove on one of the priced highway links, the distance was recorded by the GPS unit, and the user charge was displayed on a taxi meter–like device in the vehicle and deducted from the household's remaining budget. The household could keep any budget remaining at the end of the pricing period. Thus, driving on the priced links during the experiment cost them real money that they would otherwise get to keep.

In theory, the Traffic Choices data can be used to estimate disaggregate, trip-level joint models of route-type choice and departure time choice, as the price varied across link types and times of day and week. In practice, because the data are in GPS format and the study was not designed for this particular type of analysis, the team found it to be extremely challenging to use these data for choice modeling. Because the data are so potentially informative, however, and because GPS traces will be an increasingly common source of data for travel demand models, it is worthwhile to report the team's experiences and findings in analyzing these data.

The data set includes GPS traces for almost 1,000,000 auto trips in the region, and many of those are for the same individuals making trips between the same locations at the same TOD over an extended period of many months. This means that these data could be used to obtain both average speeds and travel time variability for many highway links and O-D pairs in the region. While the geographic coverage of this information is not sufficient to use in a general travel-demand model estimation for the region, it nevertheless provides a useful comparison to generalize to other network-based measures of travel time variability and reliability used in this project.

CHAPTER 3

Demand Model Specifications and Estimation Results

This chapter provides a detailed technical discussion of the main focus of the C04 research project, which was the specification and estimation of new advanced forms of travel demand models that aim to substantially improve how road pricing and congestion can be more fully and realistically modeled for transportation policy and planning.

This chapter describes the results of model estimation research in terms of the somewhat more general findings presented in two key subsections: Overview of Section, Approach, and Main Findings; and Summary Comparison and Synthesis.

The first subsection provides in-depth discussion of the conceptual and behavioral framework adopted for the C04 research and the wide range of possible responses considered to congestion and pricing. The rest of the chapter, which focuses on model estimation results developed with data from New York and Seattle, Washington, is organized in a two-dimensional fashion, with the major subsections organized by *model types* (route choice, time-of-day [TOD] choice, mode choice, and other choice dimensions) and the minor subsections organized by *model features* (main utility specification, segmentation, the incorporation of reliability, and other important model properties). The corresponding components developed and tested with the different models estimated in the course of the C04 research, using the data described in Chapter 2, are then presented back-to-back, focusing on each of the principal model choices and features proposed for improvement.

Given the complexity of the modeling issues and the need to adequately document the methods developed and applied for the C04 research for advanced modeling, Chapter 3 is necessarily technically detailed in nature. For a more thorough understanding of the topic, the technical reader is referred to an unabridged version of chapter with full detail in the discussion of all conceptual and technical details associated with the specification and estimation of the demand models addressed in this research project. The reader can find a full discussion of each model topic and data analysis in the unabridged, unedited Chapter 3 online: www.trb.org/Main/Blurbs/168141.aspx. Appendix A, which provides the full statistical results for the key models that were estimated for this study and are discussed in this chapter, is organized according to the specific types of models estimated (route choice, mode choice, and so forth) and the specific data sets used.

Structural Dimensions for Analysis of Congestion and Pricing Impacts on Demand

Possible Choice Dimensions

The behavioral framework adopted for the C04 research has been constructed to include a wide range of possible responses to congestion and pricing, organized as shown in Table 3.1 in an approximately hierarchical order from the short term to the long term.

Most of the existing models for pricing (both in research and practice) have been largely focused on the subset of trip-level short-term responses, including route, preroute, car occupancy, mode choice, and departure time choice (Brownstone et al. 2003; Brownstone and Small 2005; Lam and Small 2001; Mahmassani et al. 2005; Mastako 2003; Verhoef and Small 2004). Within this limited framework, there have been only a few examples of a full integration across all these choices: in the existing activity-based models (ABMs) developed for Columbus, Ohio (PB Consult, Inc. 2005) and Montreal, Quebec (PB Consult, Inc. 2003).

There are, however, many other important travel dimensions that have been less explored. Long-term impacts of congestion and pricing may include fundamental changes in travel behavior patterns that cannot be captured and understood at the single trip level. For example, in urban over congested areas like New York, Chicago, and San Francisco, many employers offer workers a compressed work schedule of four 10-hour days. This new choice dimension can have a significant impact on the amount of travel produced and its temporal distribution. This choice, however, is clearly not a trip-level decision

Table 3.1. Possible Responses to Congestion and Pricing

Choice Dimension	Time Scale for Modeling	Expected Impact
Network route choice	Short term—trip episode	Stratified response by user group
Preroute choice (toll versus nontoll)	Short term—trip episode	Stratified response by user group
Car occupancy	Short term—tour or trip episode	Planned and casual carpool
Mode choice	Short term—tour or trip episode	Shift to transit, especially to rail and for low- to medium-income groups
TOD or schedule	Short term—tour or trip episode	Peak spreading
Destination or stop location	Short term—tour or trip episode	Improved accessibility effect combined with negative pricing effect on trip distribution for nonwork trips
Joint travel arrangements	Short term—within day	Planned carpool or escorting
Tour frequency, sequence, and formation of trip chains	Short term—within day	Lower tour frequency and higher chaining propensity
Daily pattern type	Short term—weekly (day to day)	More compressed workdays and work from home
Usual locations and schedule for nonmandatory activities	Medium term—1 month	Compressed or chain patterns; weekly planned shopping in major outlets
Household or person mobility attributes (transponder, transit path, parking arrangements at work)	Medium term—1 to 6 months	Higher percentage of transponder users and parking arrangements for high incomes, higher percentage of transit path holders for low incomes
Household car ownership choice	Long term—1 year	Stratified response by income group: Higher car ownership for high incomes, lower car ownership for low incomes
School or university location and schedule	Long term—1 to 5 years	Choice by transit accessibility; flexible schedules
Job or usual workplace location and schedule	Long term—1 to 5 years	Local jobs for low incomes; compressed or flexible schedules
Residential location	Long term—5 years +	Income stratification: High-income suburbs around toll roads, low-income clusters around transit
Land use development	Long term—5 years +	Urban sprawl if no transit; otherwise shift to transit

comparable to a choice between managed and free lanes (or between toll and nontoll roads) for a particular trip. Choices such as this should be modeled within a proper behavioral framework, including an extended time scale, with a robust set of explanatory variables, and linkages to the other short-term and long-term choices (Pendyala 2005; Spear 2005).

In general, the important multiple possible behavioral responses that are beyond a traditional trip-level modeling of choices can be grouped into the following broad classes:

- Trip or tour destination choice that is equally important for both ABMs and four-step models. It is normally assumed that impacts of congestion and pricing should be captured through the generalized cost or mode choice logsum (Erhardt et al. 2003; Dehghani and Olsen 1999); however, there can be more direct and specific impacts that are worth exploring.
- Short-term choices that relate to daily activity patterns that cannot be fully captured at the elemental trip level. These choices include explicit joint travel arrangements (Vovsha et al. 2003; Vovsha and Petersen, 2005), tour formation (Parsons Brinckerhoff Quade and Douglas, Inc. et al. 2005), and daily pattern type (PB Consult, Inc. 2005) (e.g., the decision to stay at home on a given day). These choices can be applied only in an ABM framework (though there might be an additional use of this for four-step models in order to investigate congestion and pricing impacts on trip generation). It is important to address these dimensions along with the conventional trip dimensions because many of the new pricing forms are not trip based (e.g., daily area pricing schemes applied in London [Litman 2005] and currently envisioned or modeled in New York and San Francisco).
- Medium-term choices that relate to choice of usual location and schedule for nonmandatory activities (like shopping or entertainment). It might be beneficial for a deeper understanding and ability to forecast such choices to put certain choices into a medium-term framework in order to explore the impacts of congestion and pricing beyond a short-term single-trip consideration. This type of choice can be incorporated in an advanced ABM only.
- Medium- or long-term choices that relate to person or household mobility attributes like car ownership, transponder, transit path, and parking arrangements. There is a growing recognition of the importance of these choices in understanding and modeling impacts of congestion and pricing.

There have been some initial attempts to formulate and estimate choice models related to the acquisition of transponders (Yan and Small 2002) simultaneously with preroute, departure time, or car occupancy (or some combination of these factors), although the estimation was implemented at the single-trip level.

- Long-term location choices of residential place, workplace, and school, as well as land use development impacts. A special methodology for analysis of congestion and pricing impacts on these choices has not yet been developed. The existing long-term models of this type operate with standard trip-level measures of accessibility (Vovsha, Davidson et al. 2005); thus, the effect of a different and extended time scale is lost. The team plans to explore data sets that include information on long-term choices (along with trip records, of course) in order to ascertain the differential impacts of congestion and pricing over various time scales.

This classification of possible choice dimensions is incorporated in the formulation of a comprehensive conceptual model of travel behavior that served as the starting point in C04 for the specification of model systems that could be estimated with the selected data sets. Several of these choice dimensions represent relatively new choice models that have not yet been widely accepted and explored (only first attempts to formulate and estimate these models have been made and reported). These include integration of the binary preroute choice (toll versus nontoll) in the mode choice nested structure, payment type (cash, E-ZPass, transponder) and associated vehicle equipment, as well as models of carpooling mechanisms (explicit modeling of joint travel).

Functional Forms for Highway Utility (Generalized Cost)

As described in Chapter 1 in Highway Utility Components, the highway travel utility function is a basic expression that combines various level-of-service (LOS) and cost attributes as perceived by the highway user. It is directly used in the highway trip route choice (e.g., between the managed lanes and general-purpose lanes on the same facility). It also constitutes an essential component in mode and TOD choice utilities. The form of the highway utility function is also important for modeling other (upper-level) travel choices, as it serves as the basis for accessibility measures. Thus, it is essential to explore the highway travel utility function and its components before considering a simplified framework of route choice in the highway network, because the complexity builds when additional choice dimensions are considered.

In most travel demand models, including those developed for practical and research purposes, the highway utility function (U) takes the following simple form:

$$U = a \times T + b \times C \tag{3.1}$$

where

T = travel time;
C = travel cost;
$a < 0$ = coefficient for travel time;
$b < 0$ = coefficient for travel cost; and
a/b = value of time (VOT).

Coefficients for travel time and cost normally take negative values, reflecting the fact that travel in itself is an onerous function necessary only for visiting the activity location. Thus, the travel utility is frequently referred to as the "disutility" of travel. However, in some research, the negative character of travel utility has been questioned in some contexts. In particular, a positive travel utility was seen to be associated with long recreational trips on weekends (Stefan et al. 2007). Also, an interesting effect was observed for commuting trips, on which commuters seem to prefer a certain minimum time and are not interested in reducing it below a certain threshold (Redmond and Mokhtarian 2001).

More importantly, it is clear that the standard representation of highway travel utility as a linear function of two variables with constant coefficients is an extremely simplified one. A great deal of the C04 research effort has been devoted to the elaboration of this basic form in the following ways:

- Investigation of the highway user perception of travel time by congestion levels. This means that a simple generic coefficient for travel time could be replaced with the coefficients differentiated by congestion levels;
- Inclusion and estimation of additional components, of which travel time reliability has been currently identified as the most important. Reliability is seen as an additional and nonduplicating term along with the average travel time and cost; and
- Testing more complicated functional forms that are nonlinear in time and cost, as well as account for randomly distributed coefficients or VOT (in addition to any explicit segmentation accounting for the observed user heterogeneity). With these enhancements, VOT is no more assumed as a constant, but becomes a varying parameter depending on the absolute values of time and cost as well as reliability.

As a working model the team has adopted the following general expression for the highway travel utility that will be explored component-by-component in the current research:

$$U = \sum_{k=1}^{5} [a_k \times \varphi_k(T_k)] + \sum_{m=1}^{3} [b_m \times \phi_m(C_m)] + \sum_{n=1}^{3} c_n R_n \tag{3.2}$$

where

$k = 1$ represents the uncongested highway travel time component;
$k = 2$ represents the congested highway travel time component (extra delay);

$k = 3$ represents parking search time;
$k = 4$ represents walk access or egress time (e.g., from the parking lot to the trip destination);
$k = 5$ represents extra time associated with carpooling (picking up and dropping off passengers);
$T_k =$ (average) travel time by component;
$m = 1$ represents highway toll value;
$m = 2$ represents parking cost;
$m = 3$ represents vehicle maintenance and operating cost;
$C_m =$ travel cost value by component;
$n = 1$ represents disutility of time variation (first measure of reliability);
$n = 2$ represents schedule delay cost (second measure of reliability);
$n = 3$ represents utility of (lost) activity participation (third measure of reliability);
$R_n =$ reliability measures by component;
$a_k, b_m, c_n =$ coefficients to be estimated; and
$\varphi_k(\ldots), \phi_m(\ldots) =$ functions for nonlinear transformation of time and cost variables.

This formulation makes it more difficult to calculate VOT, although it is still possible. In the same way, value of reliability (VOR) can be calculated for the first type of reliability measure (assuming that this reliability measure is in minutes). VOR essentially represents travelers' willingness to pay for reduction in travel time variability in the same way as VOT represents their willingness to pay for (average) travel time savings. More specifically, in the context of willingness to pay tolls for saving time in congestion conditions, VOT can be calculated by the following general formula:

$$VOT(T_2, C_1) = \frac{\partial U/\partial T_2}{\partial U/\partial C_1} = \frac{a_2 \varphi_2'(T_2)}{b_1 \phi_1'(C_1)} \tag{3.3}$$

A similar calculation can be implemented for VOR. With nonlinear transformation functions, VOT and VOR are no longer simply constant values. They now vary and depend on the absolute values of time and cost variables at which the derivatives of the transformation functions are taken.

The innovative components of the C04 research that relate to perceived highway time, travel time reliability, and nonlinear transformations are discussed in the subsequent sections. It should be noted that some components, specifically perceived travel time and some reliability measures, might be correlated statistically (and are also conceptually duplicative to some extent). Thus, it is highly improbable that the entire formula (Equation 3.2) would ever be applied. Instead it serves instead as a conceptual framework in which proposed model structures can be derived and statistically tested.

Dimensions for Model Segmentation

Another long-term gap in the understanding and the modeling of congestion and pricing is associated with inadequate segmentation of population and travel. It has been generally recognized by the both research and practitioner communities that the profession needs to advance beyond the crude average VOT estimates (and other related behavioral parameters) obtained from aggregate analyses (Hensher and Goodwin 2003).

There is a significant amount of research providing insights into behavioral mechanisms and statistical evidence on the heterogeneity of highway users across different dimensions. Although income and trip purpose have been traditionally used in many models as the main factors that determine VOT, in reality VOT is a function of many other variables. In fact, in many cases, income and trip purpose might not even be the most important factors, especially when situational factors and time pressure come into play (Spear 2005; Vovsha, Davidson et al. 2005).

A variety of traveler and trip-type dimensions are understood to be important. The research team distinguishes between the following main groups:

- Socioeconomic Segments of Population. These characteristics are exogenous to all activity and travel choices that are modeled in the system. Thus, the corresponding dimensions can always be applied for any model, either for a full segmentation or as a variable in the utility function;
- Segmentation of Activities. These characteristics are exogenous to travel choices, but endogenous to activity-related choices. Thus, in the applied model system, it is necessary that the corresponding activity choices are modeled prior to the given model; otherwise they cannot be used for the model segmentation; and
- Travel Segmentation. These characteristics are endogenous to the system of travel choices. In model estimation, they have to be carefully related to the model structure to ensure that all dimensions and variables used in each particular model have been already modeled in the model chain.

The socioeconomic segmentation of population may best be addressed by the following:

- Income, Age, and Gender. A higher income is normally associated with higher VOT (Brownstone and Small 2005; Dehghani et al. 2003). Middle-age female status has also been associated with higher VOT (Mastako 2003; PB Consult, Inc. 2003);
- Worker Status. Employed persons (even when traveling for nonwork purposes), because of their tighter time constraints, are expected to exhibit a higher VOT than nonworkers; and
- Household Size and Composition. Larger households, with children, are more likely to carpool and take advantage of managed lanes (Stockton, Benz et al. 2000; Vovsha et al. 2003).

The segmentation of activities may best be addressed by the following:

- Travel Purpose. Work trips, and, in particular, business-related trips, normally are associated with higher VOT than trips for nonwork purposes (Dehghani et al. 2003; Parsons Brinckerhoff Quade and Douglas, Inc. et al. 2005; PB Consult, Inc. 2003). A frequently cited high-VOT trip purpose is a trip to the airport to catch an outbound flight (Spear 2005). A list of special trip purposes with high VOT might also include escorting passengers, visiting a place of worship, a medical appointment, and other fixed-schedule events (e.g., theater or sport events). A deeper understanding of the underlying mechanisms for such behavior would be valuable, including a combination of factors such as schedule inflexibility, low trip frequency, and situational time pressure.
- Day of Week. Weekday versus Weekend. There is statistical evidence that VOT for the same travel purpose, income group, and travel party size on weekends is systematically lower than on weekdays, including some examples of positive travel utility associated with long discretionary trips (Stefan et al. 2007). It is yet to be determined if these differences can be explained by situational variables, or if there is an inherent weekend type of behavior that is different from regular weekday behavior. In any case, whether directly or for a proxy for situational time pressure, it would be useful to test the differences statistically. A positive utility of travel has been found, most notably in the choice of distant destinations for discretionary activities on weekends (perhaps with a sightseeing or excursion component). This utility should be explored, however, to see if it is actually correlated with tolerance of congestion delays or unwillingness to pay tolls.
- Activity and Schedule Flexibility. Fixed-schedule activities are normally associated with higher VOT for trips to activity because of the associated penalty of being late; this has manifested itself in previous research that documents that VOT for the morning commute is higher than for the evening commute. Probably a similar mechanism for trips to airports (high penalty of being late) creates higher VOT estimates. The team also expects that schedule flexibility will be an important factor for nonwork activities; for example, a trip to a theater might exhibit a high VOT, but shopping might be more flexible.

The segmentation of travel can be best addressed with the following:

- Trip Frequency. Regular trips and their associated costs may receive more (or less) formal consideration than those that occur infrequently. For example, $1.50 for an auto trip to work may be perceived as $3.00/day (assuming a symmetric toll) and $60/month, thus receiving special consideration. This perceptional mechanism is likely to be very different for infrequent and irregular trips when the toll is perceived as a one-time payment. For intercity trips, travelers' recognition of the return trip is not obvious, since it may occur on a different day.
- TOD. Research confirms that a.m. and p.m. peak periods are associated with higher VOT than off-peak periods and that a.m. travelers (mostly commuters) are more sensitive to both travel time and reliability than p.m. commuters (who mostly are returning home) (Brownstone et al. 2003). However, few have explored how these phenomena relate to schedule flexibility, or how TOD factors affect VOT for nonwork trips;
- Vehicle Occupancy and Travel Party Composition. Although a higher occupancy normally is associated with higher VOT (though not necessarily in proportion to party size), it is less clear how travel party composition (e.g., a mother traveling with children, rather than household heads traveling together) affects a party's VOT.
- Trip Length or Distance. An interesting convex-shape function has been estimated for commuters' VOT (Steimetz and Brownstone 2005). For short distances, VOT is comparatively low since the travel time is insignificant, and delays are tolerable. For trip distances around 30 miles, VOT reaches a maximum; however, for longer commuters VOT goes down again, because they presumably have self-chosen residential and work places based on the long-distance travel. Additionally, in the context of mode choice, strong distance-related positive biases have been found for rail modes in the presence of congestion (as a manifestation of reliability [Parsons Brinckerhoff Quade and Douglas, Inc. et al. 2005]) and carpools (since carpools are associated with extra formation time).
- Toll Payment Method. This is an important additional dimension that has not been explored in detail. An analysis done by the Port Authority of New York and New Jersey has shown that the introduction of E-ZPass at its tolled crossings attracted a significant new wave of users despite a relatively small discount (Holguín-Veras et al. 2005). In the same way in which transportation analysts speak about perceived time, we should also probably speak about perceived value of money in the context of pricing. Bulk discounts and other nondirect pricing forms should be modeled at the daily pattern level rather than trip level. We also have to understand the impact of congestion on the whole daily pattern rather than by single trips, including analysis of daily time budgets and trade-offs made to overcome congestion (including work from home, compressed workweeks, compressed shopping, and moving activities to weekends).
- Situational Context. Time Pressure versus Flexible Time. This is recognized as probably the single most important factor determining VOT that has proven difficult to measure

and estimate explicitly, as well as to include in applied models (Spear 2005; Vovsha, Davidson et al. 2005). There is evidence that even a low-income person would be willing to pay a lot for travel time savings if he or she were in a danger of being late to a job interview or were escorting a sick child. This factor is correlated with the degree of flexibility in the activity schedule (inflexible activities, trips to airport, fixed schedules, and appointments will be the activities most associated with time pressure), but does not duplicate it. For example, for a high-income person traveling to the airport, the VOT might not be relatively low if this person has a 4-hour buffer before the departure time. With ABMs, the analyst could use the number of trips or activities implemented by the person in the course of a day, as well as the associated time window available for each trip or activity, as an instrumental proxy for time pressure.

In model formulation, estimation, and application, it is crucial to follow a conceptual system design and obey the rules of application of those variables that are exogenous to the current model. For example, if the TOD model is placed after mode and occupancy choice, then mode and occupancy can be used as the TOD model segmentation. However, TOD in this case cannot be used for segmentation of the mode and occupancy choice models. If the order of models is reversed (TOD choice before mode and occupancy choice), then the segmentation restrictions would also be reversed. When different models are estimated it is essential to keep a conceptual model system (or at least a holistic framework as described below) in mind in order to make these models compatible and avoid endogeneity–exogeneity conflicts.

It should be understood that all these dimensions cannot be simultaneously included in operational models as explicit segments in Cartesian combination. With a four-step model framework, this would immediately result in an unfeasibly large number of trip tables. The disaggregate basis of the ABM framework is more flexible, and theoretically can accommodate any number of segments. They are, however, limited in practical terms by the sample size of the travel survey (normally several thousands of individuals), which quickly wears thin for multidimensional segments. However, there are other ways to constructively address segmentation in operational models. They include flexible choice structures with parameterized probabilistic distribution for parameters of interests (e.g., VOT), as well as aggregation of segments by VOT for assignment and other model components that are especially sensitive to dimensionality.

It should also be understood that VOT represents only one possible behavioral parameter, and that it is essentially a derived one. In most model specifications and corresponding estimation schemes, VOT is not directly estimated, but rather derived either as the ratio of the time coefficient to cost coefficient (in simple linear models as specified in Equation 3.1) or as the marginal rate of substitution between time and cost (in a general case as specified in Equation 3.3). Thus, very different behaviors can be associated with the same VOT. For example, both time and cost coefficients can be doubled, which leaves VOT unchanged; however, there would be very different estimated responses to congestion and pricing in these two models. Large coefficients will make the model more sensitive to any network improvement or change in costs, whereas smaller coefficients will make it less sensitive.

One of the most detailed VOT segmentation analyses of the type described in the previous subsection was carried out for the Netherlands National Value of Time study (Bradley and Gunn 1990), which used 10 simultaneous segmentation variables. A similar approach was used for national studies in the United Kingdom and Sweden.

All else being equal, a more detailed segmentation typically tends to dampen the overall price sensitivity across the population, since a typical sigmoid response curve, like the logit model, has the steepest (most elastic) part in the middle, but the ends are quite flat, and market segmentation tends to move distinct groups away from the middle.

Measures of Travel Time Reliability

In general, four possible methodological approaches to quantifying travel-time reliability are either suggested in the research literature or already applied in operational models:

- Indirect Measure: Perceived Highway Time by Congestion Levels. This concept is based on statistical evidence that in congestion conditions, travelers perceive each minute with a certain weight (Small et al. 1999; Axhausen et al. 2007; Levinson et al. 2004; McCormick Rankin Corp. and Parsons Brinckerhoff 2008). Perceived highway time is not a direct measure of reliability because only the average travel time is considered, although it is segmented by congestion levels. It can serve, however, as a good instrumental proxy for reliability because the perceived weight of each minute spent in congestion is in part a consequence of associated unreliability.
- First Direct Measure: Time Variability (Distribution). This is considered as the most practical direct approach and has received considerable attention in recent years. This approach assumes that several independent measurements of travel time are known that allow for forming the travel time distribution and calculation of derived measures, such as buffer time (Small et al. 2005; Brownstone and Small 2005; Bogers et al. 2008). One of the important technical details with respect to the generation of travel time distributions is that even if the link-level time variations are known, it is a nontrivial task to synthesize the O-D–level time

distribution (reliability skims) because of the dependence of travel times across adjacent links due to a mutual traffic flow. The implementation challenge posed by this issue was specifically addressed in the course of this project.

- Second Direct Measure: Schedule Delay Cost. This approach has been adopted in many academic research works on individual behavior (Small 1982; Small et al. 1999). According to this concept, the direct impact of travel time unreliability is measured through cost functions (penalties in expressed in monetary terms) of being late (or early) compared with the planned schedule of the activity. This approach assumes that the desired schedule is known for each person and activity in the course of the modeled period. This assumption, however, is difficult to meet in a practical model setting.
- Third Direct Measure: Loss of Activity Participation Utility. This method can be thought of as a generalization of the schedule delay concept. It is assumed that each activity has a certain temporal utility profile and that individuals plan their schedules to achieve maximum total utility over the modeled period (e.g., the entire day) taking into account expected (average) travel times. Any deviation from the expected travel time due to unreliability can be associated with a loss of a participation in the corresponding activity (or gain if travel time proved to be shorter) (Supernak 1992; Kitamura and Supernak 1997; Tseng and Verhoef 2008). Recently this approach was adopted in several research works on dynamic traffic assignment (DTA) formulation integrated with activity scheduling analysis (Kim et al. 2006; Lam and Yin 2001). Similar to the schedule delay concept, however, this approach suffers from data requirements that are difficult to meet in practice. The added complexity of estimation or calibration of all temporal utility profiles for all possible activities and all person types is significant. This makes it unrealistic to adopt this approach as the main concept for the current project. This approach, however, can be considered in future research efforts.

The details of each approach are considered in the subsequent sections.

Perceived Travel Time Weights by Congestion Levels

Variations in the perceived utility of components of transit travel time have been long recognized and used in travel models. For example, in most mode choice models and transit assignment algorithms, out-of-vehicle transit time components like *wait* and *walk* are weighted compared with in-vehicle travel time. It is not unusual to apply weights in the range of 1.5 to 4.0, reflecting that travelers perceive out-vehicle time as more onerous than in-vehicle time.

In contrast to transit modeling practice, virtually all travel models used for highway analysis include a single generic term for highway time; that is, the same coefficient is applied for each minute of highway time regardless of travel conditions. However, there is some compelling statistical evidence that highway users perceive travel time differently by congestion levels. For example, driving in free-flow conditions is likely to be perceived less negatively than driving in heavily congested (stop-and-go) conditions. It is an intuitive and behaviorally appealing notion that highway users driving in congested conditions might perceive the longer travel time as an additional delay or penalty on top of free-flow (or some expected reasonable) time. With a segmentation of travel time coefficients by congestion levels, the time spent on links with congested conditions is expected to have a larger disutility. A larger disutility associated with congestion would have at least two behavioral interpretations:

- A negative psychological perception that is similar to the weight for walking to or waiting for transit service; and
- A simplified operational proxy for reliability that should be explored in combination with the explicit reliability measures.

Several research studies report statistical evidence of quite high perceptional weights that highway users put on travel time in congested conditions (Small et al. 1999; Axhausen et al. 2007; Levinson et al. 2004; McCormick Rankin Corp. and Parsons Brinckerhoff 2008; Wardman et al. 2009). Also, there have been multiple indications in recent analyses of travel surveys that the perception of the time saved by respondents in the revealed preference (RP) survey is about twice the actual measured time saved (Small et al. 2005; Sullivan 2000). In the RP framework, this might well be a manifestation that travelers operate with perceived travel times in which time spent traveling through congested segments is psychologically doubled.

In order to illustrate the magnitude of the possible weights, as well as possible approaches to differentiate travel time by congestion levels, three examples of estimated perceptions of travel time are discussed below. It should be noted that in both cases, the approaches are very simple to implement on the supply side. The network simulation can be performed, and the required LOS skims can be generated by static assignment methods, although DTA could offer additional benefits. This technique can be easily applied with both ABMs and four-step models.

In the first example (Small et al. 1999), travel time was broken into two parts:

- Time in uncongested conditions (LOS A to D); and
- Time in congested conditions (LOS E to F, i.e., close to the stop-and-go condition).

Table 3.2. Cost of Shifting 1 Minute from Uncongested to Congested Time

Total Travel Time (min)	Cost of Shifting 1 min from Uncongested to Congested Time ($)	Equivalent in VOT ($/h)
10	0.77	46.2
15	0.51	30.6
20	0.30	18.0
30	0.26	15.6
45	0.17	10.2
60	0.13	7.8

The choice framework presented in the stated preference (SP) survey context included only route choice. Travel time and cost variables were not estimated, but were stated in the SP questionnaires. The highway utility expression included total time, cost, and percentage of congested time. Using the previously introduced notation, the adopted utility specification can be written in the following way:

$$U = a \times (T_1 + T_2) + b \times C + c \times \frac{T_2}{T_1 + T_2} \tag{3.4}$$

This expression is different from the suggested formula (Equation 3.2), but it could be transformed into an equivalent formula with certain assumptions (fixed total travel time). The estimation results confirmed a very high significance for the additional term of percentage of congested time. The authors translated it into a recommended mark-up value of 2.5 to VOT savings under congested conditions compared with uncongested conditions. More detailed estimation results are summarized in Table 3.2. By virtue of the specified utility function, the cost of shifting 1 minute from uncongested to congested time is dependent on the total travel time. For an average time of 30 minutes, the VOT equivalent of the additional perceived burden associated with only congestion itself is about $15/hour, which is roughly equal to the average commuting VOT applied in most models.

The second example is taken from the recently completed travel demand model for the Ottawa–Gatineau, Canada, region (McCormick Rankin Corp. and Parsons Brinckerhoff 2008). The model framework, choice context, and utility formulation were different from those used in the Small et al. (1999) study. However, the bottom-line results look similar in many respects. In this study, a mode choice model was estimated for five travel purposes and two TOD periods (a.m. and p.m.) based on the RP data from the large household travel survey (23,870 households representing a 5% sample). Travel time and cost variables were provided from static assignment equilibrium skims from the modeled network.

The highway utility included travel cost with one generic coefficient and travel time broken into the following two components—note that this breakdown of travel time is different from the one adopted for Small et al.:

- Free-flow (minimal) time; and
- Extra delay, calculated as congested time minus free-flow time for the entire origin–destination (O-D) path.

The highway utility function had the following form:

$$U = a_1 \times T_1 + a_2 \times T_2 + b \times C + \sum_s (d_s \times h_s) \tag{3.5}$$

where

s = additional mode-specific constants and household or zonal variables;
h_s = values of additional variables; and
d_s = estimated coefficients.

The estimation results are shown in Table 3.3 as translated into VOT terms. They indicate that for several segments, specifically a.m. and p.m. work trips, as well as p.m. discretionary trips, each minute of congestion delay is perceived as about twice as onerous as the free-flow (minimal) time component. For other segments, however, statistical tests did not show a significant difference between free-flow and congestion time components; hence, two coefficients were pooled together.

Table 3.3. VOT Estimates for Free-Flow Time and Congestion Delay

	VOT ($/h)			
	a.m.		p.m.	
Trip Purpose	Free-Flow Time	Congestion Delay	Free-Flow Time	Congestion Delay
Work	22.2	42.7	19.4	40.0
University	10.0	10.0	11.0	11.0
School	5.1	5.1	5.1	5.1
Maintenance	10.7	10.7	12.1	12.1
Discretionary	9.0	9.0	11.4	29.3

Table 3.4. Highway Time Weight by Congestion Levels

Travel Time Conditions	United Kingdom	United States
Free flow	1.00	1.00
Busy	1.05	1.03
Light congestion	1.11	1.06
Heavy congestion	1.31	1.20
Stop and start	1.20	1.38
Gridlock	1.89	1.79

The third example is taken from the research work of Wardman et al. (2009), who provided new evidence on the variation in the valuation of motorists' travel time savings across a finer gradation of traffic-condition types (six levels of congestion) than had been previously attempted by means of analyzing SP data collected from different tolled roads in the United Kingdom and the United States. The summary of the time relativities is presented in Table 3.4. The study further supports a finding that a reasonable value for the perceived time weight in congested conditions lies in the range 1.3 to 2.0.

Mean Variance, Buffer Time, and Other Time Variability Measures

Time variability can be measured by any compact measure associated with a travel time distribution (e.g., any combination of the mean, dispersion, or higher moments). Taking into account such considerations as behavioral realism and simplicity of the model estimation (specifically, the formulation of SP alternatives) and application, three main forms have been proposed and tested to date (Batley et al. 2008):

- *Standard deviation* is a symmetric measure that assumes that being early or late is equally undesirable (probably not a realistic assumption for many trips and underlying activities);
- The difference between the 80th, 90th, or 95th and the 50th percentile (median) of travel times is frequently referred to as *buffer time.* This is an asymmetric and more behaviorally appealing measure because it specifically targets late arrivals and is less sensitive to early arrivals; and
- Simplified asymmetric measures in terms of *probability of certain delays* with delay thresholds such as 15 or 30 minutes are frequently used in the SP framework.

An illustrative example of the standard deviation approach is provided in Small et al. (1999) in the context of a binary route choice. The following form of utility function was adopted:

$$U = a \times T + b \times C + c \times SD(T) \tag{3.6}$$

where $SD(T)$ is the standard deviation of travel time.

Table 3.5. Value of Reliability Measured as Standard Deviation of Time

Trip Purpose and Income Group	Value of Reliability As SD(T)	
	($/min)	($/h)
Work trips, higher income	0.258	15.5
Work trips, lower income	0.215	12.9
Nonwork trips, higher income	0.210	12.6
Nonwork trips, lower income	0.167	10.0

Note: SD(T) = standard deviation of time.

Standard deviation of travel time was calculated based on the set of five travel times presented in the SP questionnaire for each highway route alternative. The estimation results showed that highway users assign a very high value to each minute of standard deviation, comparable with or even higher than the VOT associated with average travel time itself (i.e., $c \geq a$). In addition, a certain logical variation across trip purposes and income groups was captured as summarized in Table 3.5 (for one of several reported model specifications).

A good example of the second type of variability measure was presented by Small et al. (2005). The adopted quantitative measure of variability was the upper tail of the distribution of travel times, such as the difference between the 80th and 50th percentile travel times (see Figure 3.1). The authors argue that this measure is better than a symmetric standard deviation, because in most situations, being late is more crucial than being early, and many regular travelers will tend to build a safety margin into their departure times that will leave them an acceptably small chance of arriving late (i.e., planning for the 80th percentile travel time would mean arriving late for only 20% of the trips).

The choice context included binary route choice between the managed (tolled) lanes and general-purpose (free) lanes on a section of SR-91 in Orange County, California. The survey included actual users of the facility, and the model was estimated on the mix of RP and SP data. The variation of

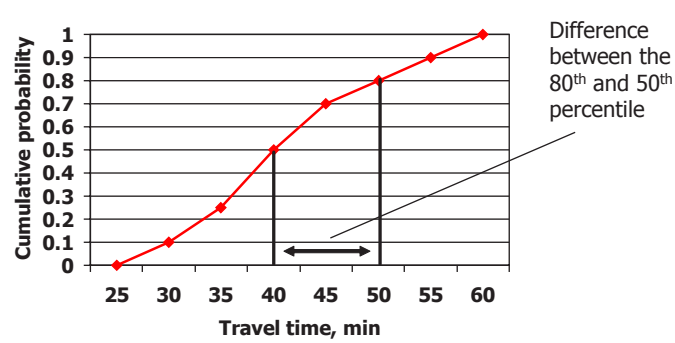

Figure 3.1. Travel time variability measure.

travel times and tolls was significantly enriched by combining RP data from actual choices with SP data from hypothetical situations that were aligned with the pricing experiment. Distribution of travel times was calculated based on the independently observed data. The measures were obtained from field measurements on SR-91 taken at many times of day on 11 days. It was assumed that this distribution was known to the travelers based on their past experience. The utility function was specified by the following formula:

$$U = a \times T + b \times C + c \times R(T) \tag{3.7}$$

where $R(T)$ is the difference between the 80th and 50th percentile.

Reliability, as defined above, proved to be valued by travelers as highly as the median travel time (VOT was roughly equal to VOR; i.e., $a \approx c$). This particular model form, with the condition of equal VOT and VOR, has a very interesting and intuitive interpretation; it could be used for a model formulation in a slightly simplified form if it were assumed from the outset that $a = c$. Indeed, if it is assumed that the willingness to pay for saving 1 minute of average travel time (the 50th percentile) is equal to the willingness to pay for 1 minute of reduction of the difference in time between the 80th and 50th percentiles, then both terms can be combined in the highway utility function because they have the same coefficient. This means that the underlying decision-making variable is the travel time value at the 80th percentile. This variable essentially combines both average travel time and time variation measure.

An example in Table 3.6 illustrates this possible approach. In the example, it is assumed that the highway user has to choose between two roads for commuting that are characterized by different time distributions. Road 1 is longer but more reliable; the travel time varies from 41 to 50 minutes. Road 2 is shorter, but travel time is less predictable and varies from 29 to 52 minutes. It is assumed that the highway user is familiar with both roads and makes his or her choice based on a rational consideration of the known distributions. In practical terms, this can be interpreted as a recollection of at least 10 trips on each road in the past, sorted by travel times from the best to worst.

Table 3.6. Illustration of Reliability Impact on Route Choice

Percentile	Travel time (min)		Preference
	Road 1	Road 2	
10	41	29	
20	42	30	
30	43	35	
40	44	39	
50	45	40	Road 2 by conventional approach
60	46	41	
70	47	45	
80	48	50	Road 1 by suggested approach
90	49	51	
100	50	52	

Although Road 2 has a better (lower) average travel time and would be preferred in most conventional modeling procedures, Road 1 has a better 80th percentile measure. In reality, the user would probably prefer Road 1 as the more reliable service. This choice framework with a single measure can be used as a simplified version of the approach. Rather than estimating two terms (average travel time and additional time associated with 80th–50th percentile), a single measure of the 80th percentile (or any other percentile larger than 50th if it yields a better statistical fit) could be used. For example, in a similar context, a 90th percentile measure was used by Brownstone and Small (2005). This framework is based on a plausible assumption that travelers under congestion conditions, characterized by travel time uncertainty, behave as rational risk minimizers. They do not base their decisions on the average values. However, they do not adopt the extreme mini–max approach (minimize risk and choose according to the worst possible case), either. The decision point probably lies somewhere between the 80th and 90th percentiles.

It is important to note that making this approach operational within the framework of regional travel models requires explicitly deriving these measures from simulation of travel time distributions, as well as adopting assumptions regarding the ways in which travelers acquire information about the uncertain situation they are about to experience. DTA and traffic microsimulation tools are crucial for the application of models that include explicit travel time variability, because static assignment can only predict average travel times.

Other approaches for measuring variability of travel time can also be considered. They are similar to the approach described above in conceptual terms, but they use a different technique in both the model estimation and the application stages. For example, in the travel model developed by PB Consult, Inc. (2003), the probability of delays longer than 15 and 30 minutes was introduced in the SP questionnaires for trucks. The subsequent estimation of the choice model revealed a very high significance of this variable that was comparable with the total trip time (in line with the VOR estimation of Small et al. [2005]). Application of this model required special probability-of-delay skims that were calculated based on the observed statistics of delays as a function of the modeled volume-to-capacity (V/C) ratio. Although this technique requires a multiday survey of travel times and speeds, it can be applied in combination with the static assignment method. Many regions with continuous traffic monitoring equipment

now have such data available for important highway segments. A problem yet to be resolved, however, is that when calculating the travel time reliability measure over the entire O-D path, the highway links cannot be considered independent.

Reliability is closely intertwined with VOT. In RP models, if variability is not measured explicitly and included as a variable, this omission will tend to inflate the estimated value of average time savings. In reality, variability in travel time tends to be correlated with the mean travel time, and people are paying for changes in both variables, so omitting one will tend to attribute the total effect to the other. Consequently, an important use of SP data sets that include reliability is to use them in combination with RP data sets for which good objective estimates of travel time variability can be derived.

It should be mentioned that the direct use of travel time variability in the framework of behavioral modeling is not the most appealing approach when compared with the other two approaches (discussed below). The principal conceptual drawback of this approach is that it does not explicitly consider the nature of underlying activities and mechanisms that create the disutility. Needless to say, the largest part of the disutility associated with unreliable travel time is being late (or too early) at the activity location, and consequently losing some part (or in some cases all) of one's participation in the planned activity. The clear practical advantage of the time variability approach, however, is in its relative simplicity and exclusive reliance on the data supplied by the transportation networks.

Schedule Delay Cost Approach

This approach has been widely accepted by the research community since its inception (Small 1982). According to this approach, the impact of travel time (un)reliability is measured by the explicit cost associated with the delayed or early arrival at the activity location. This approach considers a single trip at a time and assumes that the preferred arrival time that corresponds to zero schedule cost is known. The essence of the approach is that the trip cost (i.e., disutility) can be calculated as a combination of the following three components:

α = value of travel time and cost;
β = cost of arriving earlier than the preferred schedule; and
γ = cost of arriving later than the preferred schedule.

By definition, only one of the schedule costs can have a nonzero value in each particular case, depending on the actual arrival time versus the preferred one. There can be many analytical forms for the schedule cost as a function of the actual time difference (delay or early arrival). It is logical to assume that both functions should monotonically increase with respect to the time difference. It is also expected, in most cases, that the schedule delay function should be steeper than the early arrival function for most activities (being late is more onerous than being early). The details, however, depend on the activity type, person characteristics, and situational context.

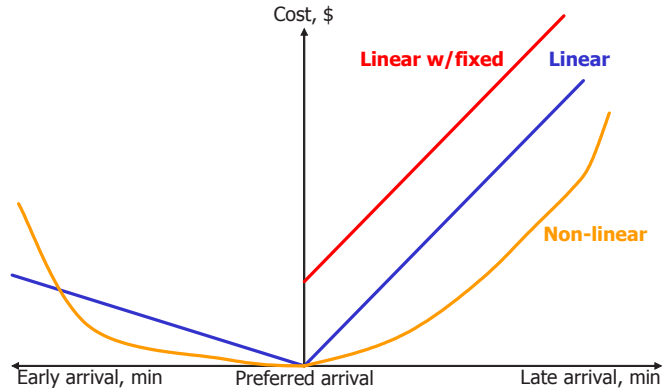

Figure 3.2. Schedule delay cost functions.

The most frequently used forms include simple linear function (i.e., constant schedule delay cost per minute), nonlinear convex function (assuming that large delays are associated with a growing cost per minute), and various piecewise functions accounting for fixed cost associated with any delay along with a variable cost per minute, as shown in Figure 3.2.

An example of a schedule delay model estimated in a highway route choice context with a specially designed SP survey is given in Small et al. (1999). The utility function was specified in the following way:

$$U = a \times T + b \times C + c \times SD(T) + \beta(\Delta t) + \gamma(\Delta t) \tag{3.8}$$

where

Δt = difference between actual and preferred arrival time;
$\beta(\Delta t)$ = early arrival cost specified as a nonlinear convex function; and
$\gamma(\Delta t)$ = late arrival cost specified as a linear function with a fixed penalty.

The estimation results with respect to the schedule delay cost are summarized in Table 3.7 for one of the tested model

Table 3.7. Estimation Example for Schedule Delay Cost

Component	Marginal Value ($)
Early Arrival (nonlinear)	
By 5 min	0.028/min
By 10 min	0.078/min
By 15 min	0.128/min
Late Arrival Dummy	
Work trips	2.87
Nonwork trips	1.80
Late Arrival (linear)	0.310/min
Extra Late Arrival Dummy	0.98

specifications. Interestingly, as reported by the authors, in the presence of explicit schedule delay cost, the travel time variability measure (standard deviation) lost its significance. The authors concluded that in models with a fully specified set of schedule costs, it is unnecessary to include the additional cost of unreliability of travel time (standard deviation).

Schedule delay cost should be distinguished from TOD choice and the associated disutility of shifting the planned (preferred) trip departure or arrival time, although in practical estimation analysis the data might mix these two factors. To clearly distinguish between the planned schedule and schedule delay, the person should explicitly report actual and preferred arrival times for each trip. Schedule delay cost assumes that the person has planned a certain schedule, but in the implementation process on the given day the delay occurs to disturb this plan. TOD choice relates to the stage of schedule planning. The outcome of this process is the preferred arrival time.

In comparing schedule delay to time variability as two different measures of time reliability, it should be noted that the schedule delay approach provides a better behavioral insight than travel time variability. It explicitly states the reasons and attempts to quantify the factors of the disutility associated with unreliable travel time, specifically perceived penalties associated with not being at the activity location on time. The schedule delay approach, however, has its own theoretical limitations as identified in the following:

- The approach is applied separately for each trip made by a person during the day, and it is assumed that the schedule delay cost for each subsequent trip is independent of the previous trip. Technically this approach is based on a fixed departure time and a preferred arrival time for each trip. In general, this is not a realistic assumption, since the activity duration requirements would create a dependence of the departure time for the next trip on the arrival time for the previous trip.
- This approach does not consider activity participation explicitly, though it makes a step toward such a consideration that the travel time variability approach ignores.
- If applied for the evaluation of user benefits from travel time savings, this approach must incorporate TOD choice (i.e., travelers' reconsideration of departure time in response to the changed congestion). Otherwise, travel time savings can result in early arrival penalties overweighting the value of saved travel time.

On the practical side, in order to be implementable, the schedule delay approach imposes several requirements that are not easy to meet, especially with conventional RP surveys:

- For each trip, in addition to the actual arrival time, the preferred arrival time should be identified. Although the preferred arrival time is generally known to the traveler (or perceived subconsciously), it is generally not observed by the modeler using RP-type data. To explore this phenomenon and estimate models that address it, the SP framework proved to be very effective, since the preferred arrival time and schedule delays can be stated in the design of alternatives. In some research, simplified assumptions about the preferred arrival time were adopted. For example, in Tseng and Verhoef (2008), the preferred arrival time was calculated as a weighted average between the actual departure time and would-be arrival time under free-flow traffic conditions.
- Application of this model for forecasting would again require input in the form of preferred arrival times. This could be accomplished either by means of external specification of the usual schedules on the activity supply side (which would probably be possible for work and fixed nonwork activities) or by means of a planned schedule model on the demand side. The latter would generate individual schedule plans (departure times) based on the optimal activity durations conditional on the average travel times. The subsequent simulation (plan implementation) model would incorporate schedule delay cost based on the simulated travel times.

Loss of Activity Participation Utility: Temporal Utility Profiles for Activity Participation

The third approach is based on a concept of time-dependent utility profile by activity type (Supernak 1992; Kitamura and Supernak 1997). Recently this approach was adopted in several research works on DTA formulation integrated with activity scheduling analysis (Kim et al. 2006; Lam and Yin 2001). The essence of this approach is that each individual has a certain temporal utility profile for each activity that is characterized by function $U(t)$. The utility profile can be estimated as a parametric or a nonparametric function of time, and time can be modeled in either continuous or discrete form. The utility profile represents an instant utility of participation in the activity at the given point of time (or during the discrete time unit that starts at the given point of time). The total utility of participation in the activity can be calculated by integrating the utility profile from the arrival time (τ) to departure time (π):

$$U(\tau, \pi) = \int_{\tau}^{\pi} u(t)\,dt \tag{3.9}$$

Simple utility profiles are independent of the activity duration. In this case, it is assumed that the marginal utility of each activity at each point of time is independent of the time already spent on this activity. This might be too simplifying an assumption, at least for certain activity types like household maintenance needs, in which the activity loses its value after the errands have been completed. More complicated utility

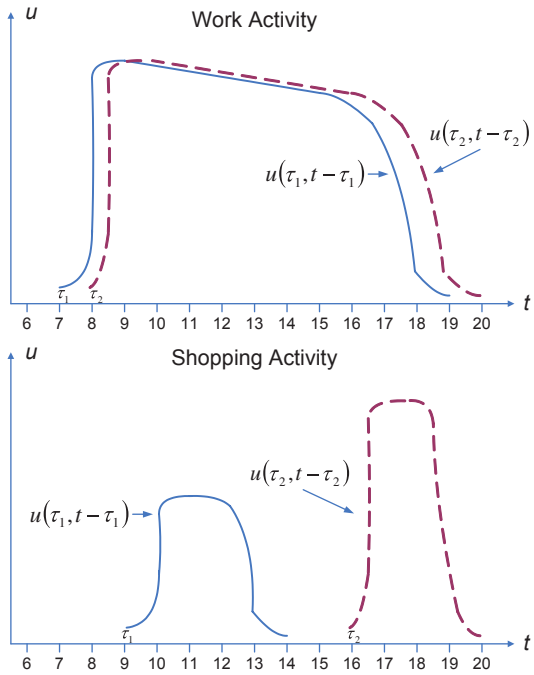

Figure 3.3. Examples of temporal profiles of activity participation utility.

profiles can be specified as two-dimensional functions $U(t,d)$, where d denotes the activity duration until moment t. In this case, the total utility of activity participation can be written as

$$U(\tau, \pi) = \int_{\tau}^{\pi} u(\tau, t-\tau)dt \tag{3.10}$$

Hypothetical but typical temporal utility profiles specified in a discrete space with an hourly resolution are shown in Figure 3.3. The work activity profile is adjusted to reflect the fixed schedule requirements (higher utility to be present at 8:00 a.m. and 5:00 p.m.). The shopping activity profile is much more uniform, with an additionally assumed convenience to undertake this activity after usual work hours. In both cases the utility is measured versus staying at home (i.e., not participating in any out-of-home activity that would require travel) as the reference (zero) utility. Thus, the utility profile can take both positive and negative values.

The concept of utility profiles is instrumental in understanding how individuals construct their daily activity schedules. According to this concept, each individual maximizes a total daily utility of activity participation. If a predetermined sequence of activity episodes is considered, it can be said that individuals switch from activity to activity when the time profile of the second activity exceeds the time profile of the previous activity. Travel episodes are placed between activity episodes in such a way that the whole individual daily schedule represents a continuous sequence of time intervals, as shown in Figure 3.4.

The effect of unreliability of travel times can be directly measured by comparing the planned and actual total daily utility of the schedule, which includes all activity and travel episodes. For simplicity, but without essential loss of generality, it is assumed that the sequence of activity episodes and trip departure times are fixed. It is also assumed that travel time delay never exceeds the planned duration of the subsequent activity; thus, activities cannot be cancelled as a result of unreliable travel time. Thus, unreliability affects only travel times and arrival times. In this context, the reliability measure can be expressed as the loss of activity participation in the following way:

$$L = \sum_i (U_i^P - U_i^A) \tag{3.11}$$

where

L = total user loss (disutility) over the whole schedule;
U_i^P = utility of the trip and subsequent activity with preferred arrival time; and
U_i^A = utility of the trip and subsequent activity with actual arrival time.

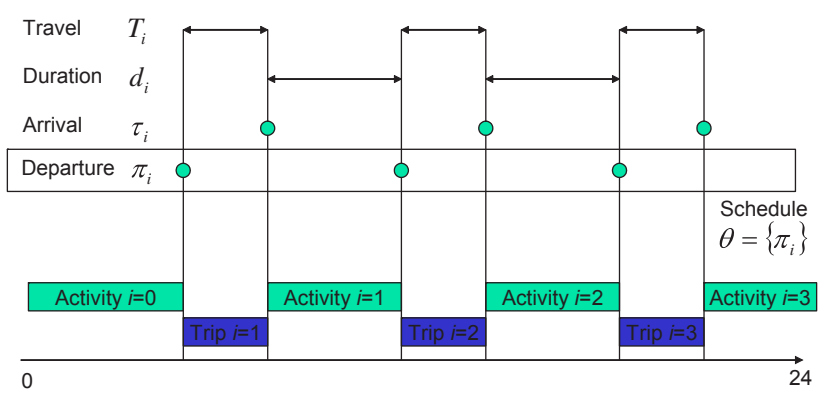

Figure 3.4. Consistent individual daily schedule.

The planned and actual utilities can be written as shown in Equations 3.12 and 3.13, respectively:

$$U_i^P(\tau_i^P) = a \times T_i^P + b \times C_i^P + \int_{\tau_i^P}^{\pi_{i+1}} U_i(t)\,dt \tag{3.12}$$

$$U_i^A(\tau_i^A) = a \times T_i^A + b \times C_i^A + \int_{\tau_i^A}^{\pi_{i+1}} U_i(t)\,dt \tag{3.13}$$

where

$$T_i^P = \tau_i^P - \pi_i;\ T_i^A = \tau_i^A - \pi_i \tag{3.14}$$

By substituting Equation 3.14 into Equations 3.12 and 3.13, and then substituting Equations 3.12 and 3.13 into the basic expression (Equation 3.11), Equation 3.15 is obtained:

$$L = \sum_i \left[a \times \left(\tau_i^P - \tau_i^A\right) + b \times \left(C_i^P - C_i^A\right) + \int_{\tau_i^P}^{\tau_i^A} U_i(t)\,dt \right] \tag{3.15}$$

where the last term (integral) represents the loss of activity participation, and the first two terms represent extra travel time and cost.

A logical relationship between temporal activity profiles of utilities and schedule delay cost was explored by Tseng and Verhoef (2008) that led to an insightful general framework. It can be shown that these two approaches are not independent. The schedule delay cost functions can always be consistently derived from the temporal utility profiles; thus, the schedule delay approach can be thought of as a particular transformation of the temporal utility profile approach. Interestingly, the opposite is true; that is, temporal utility profiles could be fully restored from the schedule delay cost functions only under some specific assumptions.

Accounting for Unobserved Heterogeneity and Situational Variability

Increasingly, travel demand analysts are looking beyond average user responses to travel costs, travel times, and other attributes toward accounting for heterogeneity or differences in user response across the population. Capturing heterogeneity in user valuation of attributes, such as travel costs and travel times, is important in order to correctly predict overall (market) responses to measures such as pricing, as well as to provide policy makers with information about the impacts of policies on different segments of the population. For example, the money VOT for users may vary considerably across a population, and policies based on the assumption of a mean money VOT may not produce the anticipated impacts (Sillano and Dios de Ortúzar 2005). As shown in the previous sections, a well-specified demand model will attempt to include as many as possible of the observable factors that can be shown to affect travel time valuation in a systematic way. However, these factors may not always be known to the analyst, and in many cases various other sources can account for the varying valuations across the population; this is referred to as *unobserved heterogeneity.*

Major advances in choice model formulation and estimation over the past decade have produced relatively robust methods to incorporate unobserved heterogeneity, particularly in the form of random coefficients (i.e., model parameters that are assumed to follow a distribution across the user population). This section provides a general framework for accounting for unobserved heterogeneity in travel demand models, specifically discrete choice models. The following subsections discuss heterogeneity, both unobserved and observed, and how to account for them within a discrete choice modeling framework. Model specification and estimation issues are briefly discussed, followed by an example to illustrate the range of questions that can be addressed with a model that accounts for unobserved heterogeneity.

Accounting for Observed and Unobserved Heterogeneity

The response of users toward attributes, such as travel time savings and cost, of different alternatives varies in general over the population of users. For example, low-income individuals are probably more concerned about and sensitive to toll prices than high-income individuals. From a practical standpoint, a common method for capturing such heterogeneity, controlling for other factors such as trip purpose, is to segment the sample of users based on exogenous criteria, such as income level, trip length, and TOD (peak versus nonpeak). Separate models are then estimated for each segment. Another practical approach is to interact attributes of the alternatives with exogenous criteria. Consider a user's choice between taking a toll or a nontoll route to work. Assume the only two attributes observed are travel cost TC_j for route type j and travel time TT_j. The importance that users place on these two attributes, reflected in the coefficients α_n and, may vary over the population, with the utility for each alternative written as

$$U_{nj} = \alpha_n TT_j + \beta_n TC_j + \varepsilon_{nj} \tag{3.16}$$

where j is either toll or free, and α_n and β_n are parameters specific to individual n. One common method for accounting for heterogeneity in response is interacting the travel time or cost terms with exogenous criteria such as income. Assuming that the importance of travel cost is inversely related to the observed income of the users (I_n), with low-income individuals placing more importance on travel costs, the coefficient for travel cost can be expressed as

$$\beta_n = \theta / I_n \tag{3.17}$$

where θ can be regarded as the mean value or importance placed on cost across all users.

An alternative approach is to allow further differences by expressing the parameters that represent preference weights (α_n and β_n) as random parameters, as opposed to point estimates, such that the distribution for these preference weights can obtained and used in the derivation of value of travel time, which in turn will be distributed across the population. These distributions can also be a function of exogenous variables. Randomness in preference weights results from a variety of reasons, possibly just because people are inherently different. Assume that users' response to travel costs, reflected by the parameter α_n, varies across all users, but is linked to unobserved factors or is intrinsically random. Examples of these unobserved or latent factors may include differences in familiarity with the network or differences in general stress levels. The resulting α_n can be expressed as

$$\alpha_n = \rho + \mu_n \tag{3.18}$$

where ρ reflects the mean response of users to travel time, and μ_n is a randomly distributed term that captures deviations from this mean value. Substituting Equation 3.17 and Equation 3.18 back into Equation 3.16 provides a utility expression that reflects both unobserved and observed heterogeneity in user responses to travel costs and travel times, as shown below:

$$\begin{aligned} U_{nj} &= \underbrace{(\rho + \mu_n)_j + (\theta / I_n) TC_j + \varepsilon\varepsilon_{nj}}_{\text{Observed or Systematic}} \\ &= \underbrace{\rho TT_j + (\theta / I_n) TC_j + \mu_n TT_j + \varepsilon_{nj}}_{\text{Unobserved or Random}} \end{aligned} \tag{3.19}$$

The above utility expression accounts for both observed and unobserved heterogeneity. The response of users toward travel costs varies systematically according to observed income, as expressed in Equation 3.17, but users' response to travel times varies randomly according to unobserved factors, as expressed in Equation 3.18. The analyst needs to specify a distribution for the random coefficient α_n. For example, the analyst may assume this distribution is normal, with mean and variance to be estimated in order to make inferences and gain insight on the distribution of users' response to travel times, including related measures such as money VOT.

Given an estimated distribution of VOT, the proportion of a population P that decides to pay a toll C_{toll} is given by the proportion with VOTs saved greater than C_{toll}:

$$P_{C_{toll}} = \int_{C_{toll}}^{\infty} (VOT) \tag{3.20}$$

The analyst selects the distribution $f(.)$ of VOT in order to find a satisfactory representation of the "true" empirical distribution. This is illustrated below in Figure 3.5, in which the proportion of payers is the blue area, given that the toll is set to $20.

The shaded area to the right is the measure of the number of people who have VOT savings exceeding the toll charged and would therefore pay it. In the case of the substantially skewed lognormal distribution, the mean is not the center of the distribution, and for the case shown in Figure 3.5, there will be fewer people in the population actually ready to pay for the toll.

The next subsection discusses the estimation in relation to discrete choice models that can capture unobserved heterogeneity and the forms these models can take.

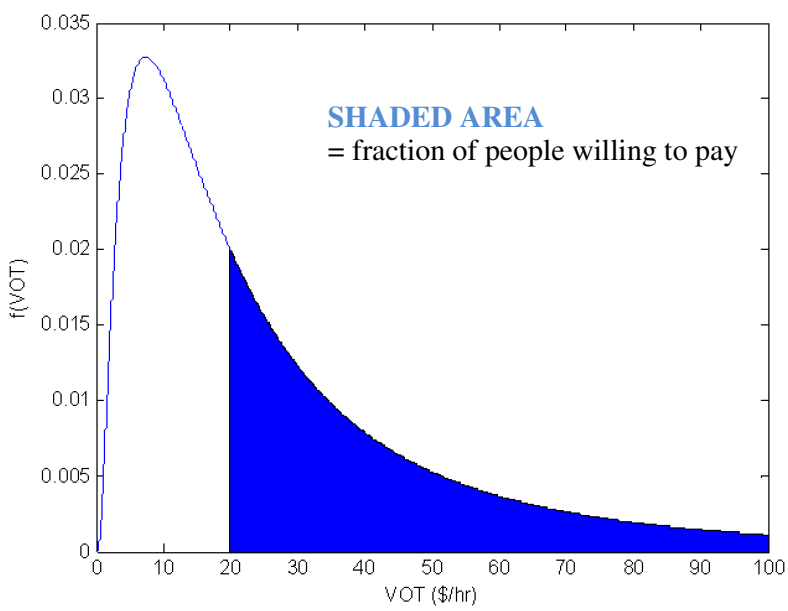

Figure 3.5. Proportion of payers with lognormal distribution for VOT for a toll of $20.

Discrete Choice Model Form and Estimation Issues

The model form these models take is dictated partly by assumptions on the error terms, which in turn are dictated by the need to account for unobserved heterogeneity. In Equation 3.19, since μ_n is not observed, the term $\mu_n TT_j$ becomes part of the unobserved component of the utility $\widetilde{\varepsilon_{nj}}$, so it can be expressed as

$$U_{nj} = \rho TT_j + (\theta / I_n) C_j + \widetilde{\varepsilon_{nj}} \tag{3.21}$$

where

$$\widetilde{\varepsilon_{nj}} = \mu_n TT_j + \varepsilon_{nj} \tag{3.22}$$

The example above illustrates the concept of heterogeneity in terms of the value individuals place on the attributes of alternatives. Heterogeneity can be captured within a discrete choice framework by linking this variation to observed or unobserved characteristics. As an example of observed heterogeneity, individual response to cost β_n was linked to an individual's income level, such that low-income individuals were more sensitive relative to high-income individuals. If individual response is linked to unobserved variables or is purely random, then the analyst would need to account for unobserved heterogeneity. In the example above, variation in response to travel time α_n was assumed to be random, as expressed in Equation 3.19, leading to total error for the utility expressed in Equation 3.22.

The type of heterogeneity present and accounted for dictates the type of choice model that is appropriate. If only observed heterogeneity is captured and accounted for, and the error term ε_{nj} is still distributed independently and identically Gumbel, a logit formulation can be used. If unobserved heterogeneity is accounted for, a logit formulation cannot be used because the total error term $\widetilde{\varepsilon_{nj}}$ is no longer distributed independently and identically. If the heterogeneity in tastes is linked to unobserved variables and is random, a logit model form would be a misspecification. As an approximation, the logit model may capture average tastes fairly well, but it cannot provide information on the distribution or heterogeneity of tastes around the average. This distribution is very important in many situations, such as forecasting the market share for tolled routes that appeal to a minority of people rather than to those of average tastes. To incorporate random taste variation appropriately and fully, probit or mixed logit model forms, or both, may be used instead.

The mixed logit and probit models are particularly well suited for incorporating unobserved heterogeneity. Continuing the previous example, assuming the coefficient for travel time varies randomly over individuals, the utility is expressed in Equation 3.18, where α_n is assumed to be distributed with a density $f(\alpha)$ with parameters θ, which can consist of a mean b and a covariance W.

The goal of estimation is to determine values for b and W. Several distributions can be assumed, both continuous and discrete. The analyst observes the travel times TT_j, but not the individual specific parameters α_n or the errors $\widetilde{\varepsilon_{nj}}$. If α_n were known, then the choice probability for an alternative, conditional on knowing α_n, would be

$$Pr_{ni}(\alpha_n) = \frac{\exp\left(\alpha_n TT_i + \left(\frac{\theta}{I_n}\right) TC_i\right)}{\sum_j \exp\left(\alpha_n TT_j + \left(\frac{\theta}{I_n}\right) TC_j\right)} \tag{3.23}$$

However, since the analyst often does not know α_n, he cannot condition on α. The unconditional choice probability is therefore the integral of $Pr_{ni}(\alpha_n)$ over all possible values of α_n:

$$Pr_{ni} = \int \left(\frac{\exp\left(\alpha_n TT_i + \left(\frac{\theta}{I_n}\right) TC_i\right)}{\sum_j \exp\left(\alpha_n TT_j + \left(\frac{\theta}{I_n}\right) TC_j\right)} \right) \cdot f(\alpha) \cdot d\alpha \tag{3.24}$$

The parameters in the choice probability in Equation 3.24 are estimated using simulated maximum likelihood estimation. This is accomplished by taking several draws from the distribution of α and averaging the choice probability Pr_{ni} across all these draws. The probabilities expressed in Equation 3.23 are approximated through simulation for given parameter values. This average simulated probability is expressed as

$$\check{Pr}_{ni} = \frac{1}{R} \sum_{r=1}^{R} Pr_{ni}(\alpha_n^r) \tag{3.25}$$

$$Pr_{ni}(\alpha_n^r) = \frac{\exp\left(\alpha_n^r TT_i + \left(\frac{\theta}{I_n}\right) TC_i\right)}{\sum_j \exp\left(\alpha_n^r TT_j + \left(\frac{\theta}{I_n}\right) TC_j\right)} \tag{3.26}$$

where R is the number of draws. The simulated probabilities (Equations 3.11 and 3.12) are inserted into the log likelihood function to give a simulated log likelihood function (SLL):

$$SLL = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \ln(Pr_{ni})$$

where $d_{nj} = 1$ if person n chose j, and $d_{nj} = 0$ otherwise. The maximum simulated likelihood estimator is the value of the parameters that maximizes SLL.

Distributions for Travel Time Coefficient

Several distributions may be assumed for the travel time coefficient β^{time}, although commonly for VOT studies, this is assumed to be a truncated normal or truncated lognormal

distribution. The normal distribution has been shown to cause some problems when applied to coefficients of undesirable attributes, such as travel time and cost, due to the possibility of positive coefficient values for these attributes (Hensher and Greene 2000; Cirillo and Axhausen 2006). To circumvent this, the normal is usually truncated to ensure that coefficients are negative for undesirable attributes and positive for desirable attributes. The lognormal distribution has the useful property of being bounded below by zero. It is useful for coefficients of attributes that are liked (or disliked) by all users. The sign is reversed for undesirable attributes, such as a travel time variable, such that the coefficient is necessarily negative. In studies of willingness to pay, the lognormal distribution has been shown to produce large and unreasonable variances and means (Hensher and Greene 2000; Hess et al. 2005). Evidence of this can be seen in the estimation results below for an "unbounded" lognormal distribution. To circumvent this, the lognormal distribution may need to be truncated to ensure reasonable means and variances.

The researcher specifies a distribution for the coefficients and estimates the parameters of that distribution. In most applications, $f(\alpha)$ is specified to be normal or lognormal:

$\alpha \sim N(b, W)$ or

$[\ln(\alpha)] \sim N(b, W)$ with parameters b and W that are estimated.

The lognormal distribution is useful when the coefficient is known to have the same sign for every decision maker, such as a travel time coefficient that is known to be negative for everyone. Triangular and uniform distributions have also been used (Hensher and Greene 2003). With the uniform density, β is distributed uniformly between $b - s$ and $b + s$, where the mean b and spread s are estimated. The triangular distribution has a positive density that starts at $b - s$, rises linearly to b, and then drops linearly to $b + s$, taking the form of a triangle. The mean b and spread s are estimated, as with the uniform distribution, but the density is peaked instead of flat. These densities have the advantage of being bounded on both sides, thereby avoiding the problem that can arise with normals and lognormals having unreasonably large coefficients for some share of decision makers.

One way around the unbounded nature of the normal and lognormal distributions is to truncate these distributions, specifying either a lower or upper bound, or both. In studies of willingness to pay, the lognormal distribution has been shown to produce large and unreasonable variances and means (Hensher and Greene 2000; Hess et al. 2005). Evidence of this can be seen in the estimated models for an "unbounded" lognormal distribution presented in Chapter 4. To circumvent this, the lognormal distribution may need to be truncated to ensure reasonable means and variances. Another alternative is to use an *Sb*-Johnson distribution, which requires specifying an upper and lower bound. The *Sb* distribution is useful for a variety of purposes. *Sb*-Johnson densities can be shaped like lognormals, but with an upper bound and with thinner tails below the bound. *Sb* densities are more flexible than lognormals: they can be shaped like a plateau with a fairly flat area between drop-offs on each side and can even be bimodal (Train and Sonnier 2004). When a lower bound other than zero is specified, the distribution is useful for an attribute that some people like and others dislike but for which there is a limit for how much the person values having or avoiding the attribute. In general, the analyst should specify a distribution that results in plausible behavior and provides good fit to the data.

Route-Type Choice: Revealed Preference Framework (New York Model)

Overview of Section, Approach, and Main Findings

Auto route choice in the highway network represents the simplest and most basic platform for understanding and modeling the behavior of highway users and their underlying generalized cost functions. Route choice is essentially a trip-level decision with no significant tour-level effects or constraints. In this choice context, it is assumed that trip origin, destination, departure time, and auto occupancy are fixed and are taken from the corresponding decisions that were modeled earlier in the model system hierarchy. Thus, the choice set consists of different highway network routes that may differ by time, cost, distance, reliability, or other measures, while the effects of person and household variables are included via interactions with route variables or as segmentation variables. This specification allows the analyst to focus on the basic form of the highway utility (generalized cost) function that is incorporated as part of the (more complicated) mode and TOD utility expressions.

Despite the attractiveness of the route choice framework as a platform for analyzing the highway utility function, there is limited supporting evidence in the literature on estimated route choice models. This is primarily due to the lack of available data sets with actual auto route itineraries reported or recorded. The common practice in most travel surveys is to collect only trip origins and destinations. Even if the actual route can be restored from some indicators on major facilities used, the identification of reasonable alternative routes to form a choice set is not a trivial issue. These problems were resolved to some extent with the two RP data sets available and extended for the current research.

The first route choice data set was based on the Household Travel Survey in the 28-county New York region, collected in

1997 and 1998. In the survey, each auto trip has an attribute of toll value paid. In the New York region most tolls are clearly associated with major facilities like bridges and tunnels around Manhattan or the New Jersey Turnpike, and specific facilities are clearly the best tolled options for certain subsets of origin and destination zone pairs. A significant number of auto trip records in the survey have origins and destinations for which both a tolled route and a free route are feasible and reasonably competitive in terms of generalized cost. Thus, it proved to be possible to from a binary route-type choice model (tolled versus free) and to support it by the corresponding set of skims for the reported TOD period, including time, cost, distance, and reliability measures (generated by the method described in Chapter 2), along with the possibility of segmentation by congestion levels and facility type. The synthetic congested travel time estimates used for this model are, of course, subject to the limitations of static traffic assignment procedures, although the simulations were implemented for each hour of the day separately.

The second data set was created from the Seattle Traffic Choices Study from 2006 (see Chapter 2), for which global positioning system (GPS) time and location data streams from travel on actual routes were available. The chosen route types were identified, and the best alternative routes were constructed to support the same binary route-type choice model (a tolled freeway route versus a nonfreeway route with a lower toll cost). Travel time distributions were calculated based on the actual average time variability across the GPS traces for each network link pair (further aggregated to O-D pair) over all weekdays in the 12-month survey period. The route-type choice framework with the Seattle data was extended to incorporate TOD (departure time dimension).

As is the case with practically all RP data sets, only the first two types of reliability measures described earlier (perceived highway time and travel time distribution) were available. Analysis of the other two reliability measures (schedule delays and temporal utility profiles) could not be supported by the available data in the survey, and no reasonable way of generating these measures synthetically was found within the research project framework.

The estimation results for these models are analyzed in the following subsections and compared with other relevant results reported in the literature. The analysis begins with the most basic linear specification. In each subsequent subsection, one specific aspect is analyzed individually in relation to the base specification. In the penultimate subsection, the best features are combined in one recommended specification of the highway generalized cost function that is the main constructive outcome of the current stage of research. This form of the generalized cost function linearly combines mean travel time, cost, and travel time reliability with a consideration of the nonlinear effects of distance, income, and car occupancy on these three main terms. This form is used as a seed construction that is further analyzed in this chapter as part of extended choice models that include mode and TOD dimensions.

In the final subsection, this specification is additionally analyzed with respect to unobserved heterogeneity when some of the coefficients were estimated as random rather than as deterministic values.

Time-of-Day Choice and Joint Time-of-Day and Route-Type Choice: Revealed Preference Framework (Seattle)

Overview of Section, Approach, and Main Findings

This section explores another primary dimension—TOD choice—in the most basic trip framework. Two data sets with different model specifications were used. The first is based on the Household Travel Survey in Seattle in 2000. This data set was used to estimate a trip TOD (departure time) model.

The second data set was created from the Seattle Traffic Choices Study from 2006, for which GPS time and location data streams from travel on actual routes were available. The chosen route types were identified, and the best alternative routes were constructed to support the same binary route-type choice model (a tolled freeway route versus a nonfreeway route with a lower toll cost), as was estimated for New York and discussed earlier in Chapter 3. Travel time distributions were calculated based on the actual average time variability across the GPS traces for each network link pair (further aggregated to O-D pair) over all weekdays in the 12-month survey period. The route-type choice framework with the Seattle data was extended to incorporate TOD (departure time dimension).

The combined route-type and TOD choice model estimated for Seattle is not equivalent to the pure route-type choice model estimated for New York. However, some comparisons across the coefficients that describe the route dimension are possible. As is the case with practically all RP data sets, only the first two types of reliability measures described earlier (perceived highway time and travel time distribution) were available. Again, the analysis begins with the most basic linear specification, and in each subsequent subsection one particular aspect of the base specification is analyzed.

The following main findings regarding TOD choice are summarized as follows:

- The coefficients for the main model variables of average time and cost proved to be in a reasonable range relative to previous studies. Extra delay variables (for time longer than 1.2 of free-flow time) proved to have an additional impact on TOD as a result of avoiding driving in congestion conditions.

- A direct measure of travel time reliability like standard deviation of travel time or standard deviation of travel time per unit distance proved to be statistically significant and performed better than more elaborate measures such as buffer time (the difference between the 90th and 50th percentiles).
- TOD choice is subject to many person and household variables. In particular, variables such as full-time versus part-time work status and income proved to have a significant effect on work schedules. Part-time workers and low-income workers have shorter activity durations compared with full-time, high-income workers. The longer work activity (tour) duration corresponds to earlier departures from home and later arrivals back home for higher incomes. Interestingly, after controlling for worker status and income, age and gender proved to have only minor impacts on work schedules. More than 80% of part-time workers are female, which can explain why the gender variable might be significant if work status is not included as a variable.
- Carpools for nonwork purposes tend to have a later schedule than drive-alone nonwork trips. The majority (about 75%) of the nonwork carpools correspond to joint travel by household members for which the schedule consolidation (especially if workers are involved) required this trip to be pushed to a later (after work) hour.

These effects are explored later in Chapter 3 in a more general framework of joint mode and TOD choice with cross comparison between the New York and Seattle regions.

The following main findings regarding route-type choice can be summarized with a special emphasis on generic impacts that proved to be common for both New York and Seattle:

- When compared with the basic specification of the New York route-type choice model, in general, the travel time coefficients across travel purpose and regions proved to be in a reasonable range (from –0.02 to –0.07), with a tendency for work purpose to have a greater coefficient than nonwork purpose. However, the VOTs obtained for New York ($19–$30/hour) are significantly higher than VOTs for Seattle ($7–$12/hour).
- In the previously discussed results for New York, travel time segmentation between arterial and local roads versus highways and freeways resulted in a statistically significant difference in coefficients. Arterial and local roads were characterized by a significantly higher (negative) coefficient than for highways and freeways. The Seattle model formulation adds an additional important facet to this analysis. The advantage of driving highways and freeways manifest itself only if a substantial portion of the overall trip can be driven on highway and freeways. If the freeway component is very small it loses its advantage, since the access to and egress from the freeway become as onerous as driving through intersections and stopping at traffic lights.
- A direct measure of travel time reliability such as standard deviation of travel time or standard deviation of travel time per unit distance proved to be statistically significant and performed better than more elaborate measures like buffer time (the difference between the 90th and 50th percentiles). However, the coefficients for standard deviation and standard deviation per unit distance obtained with the New York data were significantly larger than those obtained with the Seattle data. The corresponding reliability ratio for New York exceeded 1.0 in many cases, but it stands significantly below 1.0 for Seattle.
- These results contribute to the general observation from the multitude of previous studies that simple models are in general not easily transferable. Depending on the regional conditions, model specification, and the manner in which reliability measures were generated, the reliability ratio can range between 0.5 and 2.0. For this reason, in the final synthesis and recommendations the team does not follow either the New York model or Seattle model directly but rather considers them as somewhat extreme examples.

Basic Specification, Segmentation, and Associated Value of Time

When compared with the basic specification of the New York route-type choice model, in general, the travel time coefficients across travel purpose and regions proved to be in a reasonable range (from –0.02 to –0.07), with the tendency for work purpose to have a greater coefficient than nonwork purpose. For New York, this also resulted in an expected higher VOT for work trips compared with nonwork trips, which is also the most common result with many other models. However, for Seattle a different result was obtained, with the work VOT being lower than nonwork VOT. Also, in general, the VOT obtained for New York ($19–$30/hour) is significantly higher than VOT for Seattle ($7–$12/hour). The overall difference between the regions can be easily explained by the difference in average income (and income segmentation is not applied yet). The reversed ratio between the work and nonwork VOT in Seattle is difficult to substantiate and it may be a consequence of a relatively small subset of nonwork trips with tolled routes in the Seattle Traffic Choices Study.

Impact of Congestion Levels and Facility Type

The New York route-type choice model used a different specification from the Seattle model for the facility-type analysis. Thus, a direct comparison of the facility-type impacts between the two models is difficult. However, the analysis in both regions supported somewhat complementary results. In the previously discussed results for New York, travel time

segmentation between arterial and local roads versus highways and freeways resulted in a statistically significant difference in coefficients (at least for the nonwork travel purpose). Arterial and local roads were characterized by a significantly higher (negative) coefficient compared with highways and freeways. This implies a general user preference for highways and freeways compared with arterials and local roads, which is in line with the common consideration that intersections and traffic lights, in addition to travel time itself, are perceived negatively by drivers. It might also be tempting to interpret the higher (negative) travel time coefficient for arterials and local roads as a proxy for travel time reliability (which is the case with travel time segmentation by congestion levels). However, this is questionable because the freeway congestion levels are as significant as for arterials and local roads.

The Seattle model formulation adds an additional facet to this analysis. The advantage of driving highways and freeways manifests itself only if a substantial portion of the trip can be driven on highway and freeways. If the freeway component is very small it loses its advantage since the access to and egress from the freeway become as onerous as driving through intersections and stopping at traffic lights. This finding is behaviorally appealing. In general, the team believes that further research should be encouraged with respect to segmentation by facility type and the construction of a route utility function that includes variables like facility type, intersection type, and presence of traffic lights in addition to travel time and cost.

With the New York route-type choice model discussed above, a significant differentiation of time by congestion levels was found. It was technically implemented by dividing the total auto time into free-flow time and congestion delay. It is only slightly different from the Seattle RP formulation, in which the time breakdown point was 1.2 rather than 1.0 of the free-flow time. In the Seattle RP formulation this segmentation did not work directly in the trip departure-time choice context, but the delay variable proved to be statistically significant as a shift variable. This means that highway users not only tend to avoid routes with higher congestion levels, but also tend to adjust their schedule to avoid driving in the congestion periods. However, these effects may only be proxies for direct impacts of reliability measures.

Incorporation of Travel Time Reliability Measures and Value of Reliability Estimation

Overall, the Seattle results with the Traffic Choices Study confirm the main findings described for the New York model in that standard deviation of time and standard deviation of time per unit distance performed better than other (more elaborate) measures of travel time reliability, such as a buffer time (difference between the 90th percentile and median). Standard deviation of travel time per unit distance has a significant practical advantage over a simple unscaled standard deviation because the latter is frequently correlated with the mean travel time. This is not a conceptual advantage per se, but it is a significant practical constraint that is difficult to resolve in the RP setting. (This constraint can be overcome in the SP setting, however, by controlling the input LOS data.)

The coefficients for standard deviation and standard deviation per unit distance obtained with the New York data were significantly greater in magnitude than those obtained with the Seattle data. The corresponding reliability ratio for New York exceeded 1.0 in many cases, but the reliability ratio stands significantly below 1.0 for Seattle. These results contribute to the general observation from the multitude of previous studies that simple models are in general not easily transferable. Depending on the regional conditions, model specification, and the way the reliability measures were generated, the reliability ratio can be between 0.5 and 2.0, or even exceed these limits for some particular cases (Li et al. 2010; Concas and Kolpakov 2009). For this reason, in the final synthesis and recommendations the team does not follow either the New York model or Seattle model directly, but rather considers them as somewhat extreme examples. New York is characterized by extremely high congestion levels and notoriously unpredictable travel times. Coupling this with a relatively short average travel distance for auto trips (the majority of long-distance commuters in New York use transit), a reliability ratio greater than 1.0 is behaviorally justified. Seattle has generally lower congestion levels across the region; hence, the entire unreliability scale is set differently versus the average travel time.

Impact of Gender, Age, and Other Person Characteristics

Due to the data limitations of the Seattle Traffic Choices Study, it was impossible to directly compare the results with New York in the route-type choice context. However, it should be noted that even with the New York data, for which a rich set of person and household variables was available, only some gender effects, in the form of additional toll-averse bias, proved to be statistically significant. Gender, age, worker status, and other person characteristics manifested strongly in TOD choice. The cross comparisons between New York and Seattle with respect to TOD choice are discussed with a full specification of joint mode and TOD choice model.

Impact of Income

Again, the limitations of the Seattle Traffic Choices data set prevented direct comparisons with the New York analysis. With the New York model, as discussed above, the team substantiated a general functional form of highway generalized cost for which the cost variable was scaled down by income

powered by 0.6. This formulation will be further tested in the extended choice frameworks of mode and TOD choice. Income also has a strong direct impact on TOD choice. The cross comparisons between New York and Seattle with respect to TOD choice are discussed below, with a full specification of joint mode and TOD choice model.

Impact of Car Occupancy

Again, the limitations of the Seattle Traffic Choices data set prevented direct comparisons with the New York analysis. With the New York model, as previously discussed, the research team substantiated a general functional form of highway generalized cost for which the cost variable was scaled down by car occupancy powered by 0.6. This formulation was further tested in extended choice frameworks of mode and TOD choice. Car occupancy (and joint household travel) also has a direct impact on TOD choice. The cross comparisons between New York and Seattle with respect to TOD choice are discussed below, with a full specification of joint mode and TOD choice model.

Nonlinear Level of Service and Trip-Length Effects

The previously discussed analysis with the New York data substantiated a seed functional form for an interactive term between auto time and distance for work trips. This form results in a parabolic function for VOT in which the maximum VOT is associated with a commuting distance of about 30 miles; for shorter and longer trips, VOT is reduced. An attempt to replicate this effect with the Seattle data resulted in somewhat inconclusive functional forms, with the key coefficients being statistically insignificant. Part of the problem was that the Seattle data, unlike the New York data, did not provide a sufficiently large set of long travel (commuting) distances. Although the average commuting distance in the New York metropolitan region is relatively short (7.5 miles), the household survey of 11,000 households provided a significant number of observations of commuting distances beyond 30 miles. Thus, this particular model component needed further exploration and cross-regional comparisons in the mode and TOD choice frameworks, as discussed below.

Mode and Car Occupancy Choice: Revealed Preference Framework

Overview of Section, Approach, and Main Findings

The models of mode and car occupancy choice represent the next tier of statistical analysis in which the highway travel utility (generalized cost) is considered in the multimodal context. All aspects described above for route choice are also relevant for mode choice, as well, because the highway modes and route types represent alternatives in mode choice. However, because the choice framework is substantially extended to include transit modes, there are many more potential impacts, factors, and variables that come into play. Also, the mode choice framework naturally includes a much wider set of travelers, including transit users who may have very different perceptions of travel time, cost, and reliability. Additionally, the mode choice models estimated for New York used in this section are tour based, which means that two-directional LOS variables are considered (for the corresponding outbound and inbound TOD periods). A tour framework is essential for analyzing mode preferences, since many mode constraints and relative advantages of different modes cannot be seen at the level of a single trip.

A central research question at this stage is whether the main findings regarding the functional form of highway travel utility from the route choice analysis described above would hold in the more general framework of tour mode choice. In the subsections that follow, the team applies the same approach as for the previously discussed route choice. Each major factor and its associated impacts are analyzed one at a time and are progressively incorporated into the final model structure. Each factor is statistically tested with the New York data and Seattle data, while trying to keep the model structures as close and compatible as possible. Each subsection concludes with a synthesis of main findings that proved to be common for both regions.

The following main findings regarding mode choice can be summarized with a special emphasis on the generic impacts that proved to be common for both New York and Seattle and were also similar to the route choice and mode choice frameworks:

- Both mode choice models have a rich set of explanatory variables, including LOS variables, as well as various person and household variables. The overall scale of time and cost coefficients (specifically for auto time that is in the focus of the current study) is reasonable. It should be taken into account that the LOS variables in a tour model should be approximately doubled when compared with a trip mode choice model. Thus, the corresponding coefficients for time and cost need to be halved for a trip mode choice model when compared directly with a tour mode choice model. This is the case for auto in-vehicle time; for example, for work-related travel, it is –0.014 for the New York tour mode model and –0.029 for the Seattle trip mode model. For work tours in New York and work trips in Seattle, the base model specifications showed a relatively low VOT for auto users of \$6–\$7/hour. This value is not recommended for use in other models. However, the team decided not to enforce a more

reasonable VOT at this stage, but rather to continue testing of more elaborate forms for generalized cost. For nonwork travel, the VOT values are more reasonable, although there was a significant difference between New York ($6/hour) and Seattle ($11/hour). This can be explained by the model specification differences.

- Segmentation of travel time by congestion levels brought very different results. With the New York data a statistically significant effect was confirmed and actually manifested itself in the mode choice framework much more strongly than in the route-type choice framework. The congestion delay component of travel time proved to be weighted 1.8 to 3.5 versus the free-flow time. A similar test with the Seattle data did not bring reasonable results. It may be concluded that travel time segmentation by congestion levels works better in extremely congested areas, but is questionable for less-congested regions.
- With respect to direct reliability measures, the most promising model estimated with the New York data is the model for nonwork tours in which a standard deviation of travel time per unit distance was used. The corresponding reliability ratio is about 1.5 at a 10-mile distance. The most promising models estimated with the Seattle data included a formulation with buffer time per unit distance for work and nonwork trips, although a formulation with standard deviation of travel time per unit distance for nonwork trips had the right sign for all LOS variables.
- The main common effects that relate to the impact of car ownership on mode choice can be summarized as follows. There is a common tendency for carpooling to be negatively correlated with car sufficiency. Bigger households (in terms of number of workers and in terms of overall size) with fewer cars are the most frequent carpoolers. For a subchoice between transit modes, zero-car households are logically characterized by a strong propensity to walk to transit rather than drive to transit access. The probability of walk to transit is highly affected by absence of cars, or low car sufficiency, in the household. Households from these categories constitute the majority of transit users; many of them are transit captives since they either do not have cars at all, or have fewer cars than workers; hence, at least some of them become transit captives.
- Several different approaches to account for income were explored with both models, including scaling the cost variable by income (powered by a scaling parameter that should be between zero and one) and segmentation of the cost variable coefficient by income group. Although in many cases segmentation by income group resulted in better likelihood values, the team believes that the income-scaling version is more behaviorally appealing. With the New York model, a scaling parameter value of 0.8 was established for work tours and 0.6 for nonwork tours, which is in line with the previously discussed findings for route-type choice (0.6 and 0.5, respectively). The fact that the VOT elasticity with respect to income proved to be somewhat higher in the mode choice framework compared with route-type choice framework can be explained. The mode choice framework includes transit users who in general have a lower VOT and income. The corresponding version of the Seattle model, with the coefficient values corresponding to the New York route-type choice model, justified the specification with all coefficients having the right sign and being statistically significant.
- Several alternative specifications were tried with both the New York and Seattle data to capture the best cost-sharing mechanisms for carpools statistically. They included cost scaling by the powered occupancy, as well as occupancy-specific cost coefficient. The scaling strategy prevailed in New York; segmentation of the cost coefficient by occupancy was less successful. The scaling values of 0.8 for work tours and 0.7 for nonwork tours were eventually adopted for New York because those values are in line with the route choice findings. The results for Seattle indicate that the cost sharing reflected in the Seattle RP data is perhaps less strong than in the New York data.
- With the New York data set, a dummy variable that represents person status categorized by three major types (worker, adult nonworker, and child) proved to be statistically significant and was included in the base model specification described above for nonwork travel. A richer set of behavioral impacts with respect to person characteristics was found with the Seattle model specification, including some related effects on VOT of gender, age, and part-time worker status. In this regard, the New York model and Seattle model provide complementary examples of specifications that can be combined and hybridized in many ways.
- With the New York model, the shape of the distance-effect curves, which were similar to the shape obtained for work trips in the route-type choice framework, was statistically confirmed in the more general mode choice framework. Depending on the highest order of polynomial function used in the model specification (squared or cubed), the inverted U effect can be less or more prominent, with a very small impact on the overall model fit. The explanation given above can be reiterated for the same effect in the route choice framework; that is, the lower VOT for long-distance commuters is a manifestation of restructuring the daily activity–travel pattern. The team obtained roughly the same shape for both home-based work trips (HBW) and home-based other trips (HBO) with the Seattle model, with VOT rising to a maximum at a distance of about 25 miles and then decreasing, but the effect is much more pronounced for HBO. For HBO, the maximum VOT is about twice as high as the VOT for very short trips, but for

HBW, the maximum VOT is only about 20% higher than for very short trips.

- It is important to account for the main land use and density effects in the mode choice framework to ensure a reasonable background for analysis of LOS impacts and to separate these effects from the pure effects of travel time, cost, and reliability. In the New York regional conditions, the primary effects were found by segmenting trips to and from Manhattan (strongly dominated by transit) and internal trips within Manhattan (dominated by transit, walk, and taxi). These effects were captured by stratified mode-specific constants without an impact on VOT. The Seattle data indicate a somewhat similar effect for trips to the central business district (CBD).

Basic Specification, Segmentation, and Associated Value of Time

Both mode choice models have a rich set of explanatory variables including LOS variables, as well as various person and household variables. This provides a reasonable background for further tests with different functional forms for the generalized cost. The overall scale of time and cost coefficients (specifically for auto time, which is the focus of the current study) is reasonable. It must be taken into account that the LOS variables in a tour model should be approximately doubled when compared with a trip mode choice model. Thus, the corresponding coefficients for time and cost should be halved for a trip mode choice model when it is directly compared with a tour mode choice model. This is the case for auto in-vehicle time; for example, for work-relate travel, it is –0.014 for the New York tour mode model and –0.029 for the Seattle trip mode model.

VOT is directly comparable between tour and trip models. For work tours in New York and work trips in Seattle, the base model specifications showed a relatively low VOT for auto users of \$6–\$7/hour. This value is not, however, recommended for use in other models. The team decided not to enforce a more reasonable VOT at this stage but rather to continue testing of more elaborate forms for generalized cost. For nonwork travel, the VOT values are more reasonable, although there is a significant difference between New York (\$6/hour) and Seattle (\$11/hour). This can be explained by the model specification differences. Although the New York model has generic time coefficients, cost coefficients, and VOTs, the Seattle model explicitly distinguishes between auto users and transit users by employing mode-specific time and cost coefficients. This distinction must be taken with caution, because in the choice framework, utilities are not directly associated with mode users. In fact, every traveler is exposed to all modes. However, in reality, many auto users and transit users are repetitive in their choices. Thus, the chosen modes create a latent segmentation of the users themselves, which is partially captured by the estimated mode-specific coefficients.

Travel Time Segmentation by Congestion Levels and Facility Type

Segmentation of travel time by congestion levels brought very different results in the two regions. With the New York data, a statistically significant effect was confirmed and actually manifested itself in the mode choice framework, much more strongly than in the route-type choice framework. The congestion delay component of travel time proved to be weighted 1.8 to 3.5 versus the free-flow time. It is logical that mode choice framework provides a better statistical support for this phenomenon compared with route-type choice framework. In the New York region, transit share for trips to and from Manhattan constitutes 80%, but it is less than 10% for the rest of the region, and the corresponding auto trips have the biggest congestion delay. Thus, in the mode choice framework a congestion-averse attitude of transit users in addition to auto users is captured.

A similar test with the Seattle data did not bring such reasonable results. It should be noted that the Seattle model operates with a different segmentation of time than the New York model. In the Seattle model, links are broken into two categories (overcongested and other), but in the New York model, entire trip travel time is broken into a free-flow component and congestion delay. However, in both specifications the same phenomenon is captured, and in general, the trips with a greater number of links with $V/C > 1.2$ should have the biggest congestion delay. Thus, although the two segmentation schemes are not equal, the results should be strongly correlated. The team believes that the failure of this particular component with the Seattle data is the consequence of very different regional conditions compared with New York. It may be concluded that travel time segmentation by congestion levels works well in extremely congested areas, but is questionable for less congested regions where the differences between different trips in terms of congestion are somewhat blurred by the crudeness of synthetic skims.

As mentioned above in the route-type choice context, the team does not propose this method as the main vehicle for the current research, despite the strong statistical evidence from the New York data. In general, highway travel time segmentation is only a proxy for direct measures of travel time reliability.

Incorporation of Travel Time Reliability and Value of Reliability

Introduction of direct reliability measures in both models proved to be difficult, and many attempted specifications in the models failed to produce reasonable and statistically significant results. In general, it was difficult to simultaneously obtain the

right (negative) sign on average travel time, cost, and travel time reliability measures. This model specification is inherently fragile with RP data because of the correlation between all three variables (although using a standard deviation per unit distance significantly alleviates this problem). As mentioned above, part of the problem is the synthetic nature of the reliability measures and the quality of the other LOS skims. However, with some particular specifications, it proved possible to generate a logical model structure with all three variables in place.

The most promising model estimated with the New York data is the model for nonwork tours in which a standard deviation of travel time per unit distance was used. The corresponding reliability ratio is about 1.5 at a 10-mile distance. The most promising models estimated with the Seattle data included a formulation with buffer time per unit distance for work and nonwork trips, although a formulation with standard deviation of travel time per unit distance for nonwork trips had the right sign for all LOS variables. At this stage the decision was made to continue with the most promising specifications and to explore additional effects and impacts that could interact with the impacts of LOS variables.

Impact of Household Car Availability

The rich set of explanatory variables in the New York and Seattle models results in many logical impacts of congestion and pricing on mode choice. The main common effects that relate to the impact of car ownership on mode choice with respect to auto and transit modes can be summarized as follows:

- For both work and nonwork travel, there is a common tendency for carpooling to be negatively correlated with car sufficiency. Bigger households (both in terms of number of workers and overall size) with fewer cars are the most frequent carpoolers. It is important to note that about 80% of the observed carpools are intrahousehold in both regions;
- For both work and nonwork travel, drive to transit requires a car. Thus, for a subchoice between transit modes, zero-car households are logically characterized by a strong propensity to use walk to transit rather than drive to transit access; and
- For both work and nonwork travel, walk to transit is highly related to the absence of cars or low car sufficiency. Households from these categories provide the majority of transit users. Many of these transit users are transit captives because they either do not have cars at all or have fewer cars than workers; hence, at least some of them become transit captives.

The New York model provides some interesting behavioral insights about using taxis, which is a very frequently used mode in Manhattan. However, taxi is not a frequent mode in Seattle, and it was not included in the Seattle model formulation. The Seattle model includes nonmotorized modes and provides some additional insights with regard to them. The New York mode choice model includes only motorized modes, because the split between motorized and nonmotorized travel is modeled in the New York Metropolitan Transportation Council's best practice model by a separate model that was added to the New York model system due to a large proportion of walk trips in Manhattan. It was found that a mode choice framework that includes motorized and nonmotorized modes is less effective in the extreme conditions of New York, where many trips are generated as nonmotorized and are not subject to mode choice per se. However, these special modes are not in the focus of the current research.

Impact of Household or Person Income

Several approaches were explored with models in both regions, including scaling the cost variable by income and segmentation of the cost variable coefficient by income group. Although segmentation by income group resulted in many cases in better likelihood values (as illustrated with the Seattle model), the team believes that the income-scaling version is more behaviorally appealing. Segmentation by income group requires an arbitrary setting of income categories that can be quite broad. Also, it does not guarantee a smooth monotonic effect across all categories.

With the New York model, scaling parameter values of 0.8 and 0.6 were established for work tours and nonwork tours, respectively. These values are in line with the previously discussed findings for route-type choice (0.6 and 0.5, respectively), although they are not identical. The fact that VOT elasticity with respect to income proved to be somewhat higher in the mode choice framework than the route-type choice framework can be explained. The mode choice framework includes transit users, who in general have a lower VOT and income. Thus, with the generic specification of the LOS variables and cost-scaling parameters, this might result in the higher sensitivity to cost. This means that the constant elasticity to cost across a wide range of income groups and modes is still an analytically convenient simplification, and some more elaborate cost-scaling forms should be explored.

The corresponding version of the Seattle model, with the coefficient values corresponding to the New York route-type choice model, justified the specification, with all coefficients having the right sign and being statistically significant. Thus, this scaling strategy for income was adopted as the main approach in the further statistical tests. As mentioned above, this functional form is also consistent with the prevailing view on VOT elasticity with respect to income (Abrantes and Wardman 2011; Borjesson et al. 2012). This formula corresponds to the constant elasticity with the coefficient less than one (i.e., weaker than a linear function).

Impact of Joint Travel and Car Occupancy

Several alternative specifications were tried with both the New York and Seattle data in order to capture the best cost-sharing coefficient statistically. They included cost scaling by the powered occupancy as well as an occupancy-specific cost coefficient. The scaling strategy prevailed in New York; segmentation of the cost coefficient by occupancy was less successful. The values of 0.8 for work tours and 0.7 for nonwork tours were eventually adopted for New York because they are in line with the route choice findings.

The results for Seattle indicate that the cost sharing reflected in the Seattle RP data is perhaps less strong than in the New York data. Similar to the results for income, the exponents appear somewhat too high to fit the Seattle data, leaving one to question whether lower values would be more appropriate or whether other changes to the model specification could be found to better match the sensitivity found in the New York data.

As discussed above in the corresponding section on the impact of car occupancy on route-type choice, some additional dimensions within this effect could be further explored. In particular, intrahousehold and interhousehold carpools can have different cost-sharing mechanisms. It is expected that cost sharing should be higher for interhousehold carpools (that means the power coefficient close to 1.0) and lower for intrahousehold carpools (power coefficient close to zero). Additionally, cost sharing between adults might be stronger than between adults and children.

Impact of Gender, Age, and Other Person Characteristics

With the New York data set, a dummy variable representing person status, categorized by three major types (worker, adult nonworker, and child), proved to be statistically significant and was included in the base model specification described above for nonwork travel. Workers are characterized by a higher propensity to use household cars for solo driving, except for commuter rail, which workers use more frequently (as it is also the commuting mode for many of them). Nonworkers carpool and use transit more frequently. Children are the most frequent carpool and taxi passengers compared with both workers and nonworkers.

A richer set of behavioral impacts with respect to person characteristics was found with the Seattle model specification, including some related effects on VOT of gender, age, and part-time worker status. For both HBW and HBO, females have a marginally significant negative coefficient, which when added to the main travel time coefficient results in VOT for females about 20% to 25% higher than for males. As age increases from a base of 18 years old, a slight increase is seen in VOT for HBW trips, and a slight decrease is seen in VOT for nonwork trips. For part-time workers, a somewhat lower VOT for HBW trips is seen, with a magnitude similar to the gender effect, but in the opposite direction and less significant.

In this regard, the New York model and Seattle model provide complementary examples of specifications that can be combined and hybridized in many ways. The best existing specifications were preserved for the further testing. It is difficult to recommend one particular model structure that would fit all possible regional conditions. Person variables can be effectively used to stratify mode choice constants (as the New York model has shown) or to stratify VOT (as the Seattle model has shown), or both.

Effect of Tour or Trip Length

With the New York model, the shape of the distance-effect curves was estimated to be similar to the shape obtained for work trips in the route-type choice framework and was statistically confirmed in the more general mode choice framework. Depending on the highest order of polynomial function used in the model specification (squared or cubed), the inverted U effect can be less or more prominent, with a very small impact on the overall model fit. The explanation given above for the same effect in the route choice framework can be reiterated, that the lower VOT for long-distance commuters is a manifestation of restructuring the daily activity–travel pattern. Long-distance commuters tend to simplify their patterns and not to have many additional out-of-home activities on the day of regular commute because the work activity and commuting take the lion's share of the daily schedule. To compensate for this, commuters tend to have compressed work weeks or telecommute more frequently, which provides an opportunity to combine nonwork activities in one particular day of the week (most frequently, Friday) when they do not commute to work.

The team obtained roughly the same shape for both HBW and HBO trips with the Seattle model, with VOT rising to a maximum at a distance of about 25 miles and then decreasing, but the effect is much more pronounced for HBO. For HBO, the maximum VOT is about twice as high as the VOT for very short trips, but for HBW, the maximum VOT is only about 20% higher than for very short trips.

Given the consistent statistical evidence from both regions with respect to work travel, the polynomial (quadratic) form was adopted as the main structure for the subsequent tests. Adoption of the same form for nonwork travel was problematic given its failure with the New York data.

Impact of Urban Density and Land Use

This component is somewhat peripheral to the main purpose of the current research. However, it was important to account for the main land use and density effects in the mode choice

framework to ensure a reasonable background for the analysis of LOS impacts and to separate these effects from the pure effects of travel time, cost, and reliability. In the New York region, the primary effects were found by segmenting trips to and from Manhattan (strongly dominated by transit) and internal trips within Manhattan (dominated by transit, walk, and taxi). These effects were captured by stratified mode-specific constants without an impact on VOT. The Seattle data indicate a somewhat similar effect for trips to the CBD, but this region does not have a metropolitan core comparable to Manhattan; thus further analysis for internal trips within the CBD made little sense.

Joint Mode and Time-of-Day Choice: Revealed Preference Framework

Overview of Section, Approach, and Main Findings

Linkage Between Mode and TOD Choice

In practice, mode choice models are often estimated separately from TOD choice models. Typically, mode choice models are estimated using the auto and transit LOS variables for the actual chosen TOD. In model application, the mode choice model can be applied conditional on the choice from the TOD choice model and, ideally, mode choice logsums for each TOD alternative are passed up to the TOD model, as well. Several of the ABM systems in use have applied this approach. Alternatively, a TOD outcome can be drawn stochastically from the aggregate shares before the application of the mode choice model, and then the TOD model can be applied conditional on the prediction from the mode choice model, predicting the TOD using the LOS for the predicted mode. The ABMs used in Denver and Sacramento use this latter approach. As it is not obvious whether TOD should be predicted conditional on mode choice or vice versa, the best approach is to estimate joint TOD and mode choice models and empirically investigate nesting structures between mode and TOD. That is the approach used for this research, with the results described in this section.

Both this model structure and the way in which utility components related to TOD choice were formed are discussed step by step below.

Seed Hybrid Time-Of-Day Choice–Duration Structure

The seed structure used in this research with the New York data is a model for scheduling travel tours that can predict departure-from-home and arrival-back-home time for each tour with enhanced temporal resolution. The model formulation is fully consistent with the tour-based modeling paradigm and is designed for application within an individual microsimulation framework. TOD choice models of this type have been estimated and applied as a part of the activity-based travel demand model system developed in the regions in and around Columbus, Ohio; Atlanta, Georgia; and Sacramento, San Diego, and the San Francisco Bay Area, California.

The model is essentially a discrete choice construct that operates with tour departure-from-home and arrival-back-home time combinations as alternatives (Vovsha and Bradley 2004). The utility structure, which is based on "continuous shift" variables, represents an analytical hybrid that combines the advantages of a discrete choice structure (flexible in specification and easy to estimate and apply) with the advantages of a duration model (parsimonious structure with a few parameters that support any level of temporal resolution, including continuous time). The model is applied with a temporal resolution of 1 hour. It is expressed in 20 alternatives for departure and arrival times from 5:00 a.m. through 11:00 p.m. as follows:

1. Earlier than 5:00 a.m.
2. 5:00–5:59 a.m.
3. 6:00–6:59 a.m.
4. 7:00–7:59 a.m.
5. 8:00–8:59 a.m.
6. 9:00–9:59 a.m.
7. 10:00–10:59 a.m.
8. 11:00–11:59 a.m.
9. 12:00–12:59 p.m.
10. 1:00–1:59 p.m.
11. 2:00–2:59 p.m.
12. 3:00–3:59 p.m.
13. 4:00–4:59 p.m.
14. 5:00–5:59 p.m.
15. 6:00–6:59 p.m.
16. 7:00–7:59 p.m.
17. 8:00–8:59 p.m.
18. 9:00–9:59 p.m.
19. 10:00–10:59 p.m.
20. 11:00 p.m. or later.

This is expressed in $(20 \times 21)/2 = 210$ hour-by-hour departure–arrival time alternatives. Only feasible combinations in which the arrival hour is equal to or later than the departure hour are considered.

Analogue Between Discrete Choice and Duration Models Through Shift Variables

Consider a discrete set of time-related alternatives, such as alternative duration for some activity in hours $t = 1, 2, 3, \ldots, n$.

A general form for the probabilistic model that returns the probability of activity duration is

$$P(t) = f(t) \tag{3.27}$$

where $f(t)$ represents a probability density function for duration. This general form is not really operational because it incorporates any possible parametric or nonparametric density function and does not suggest any constructive method for model estimation.

Duration models operate with a special function $0 < \lambda(t) < 1$ that represents a termination rate (frequently called *hazard* in the literature) at time t assuming that the activity has not been terminated before (i.e., at one of the time points $1, 2, \ldots, t-1$). The probability density function for a duration model in discrete space takes the following form:

$$P(t) = \lambda(t)\prod_{s=1}^{t-1}[1-\lambda(s)] \tag{3.28}$$

There is a direct correspondence between the general-form density function and the continuous duration model. Any duration model has the correspondent density function calculated by Equation 3.28, and any density function has the underlying termination rate calculated by the following formula:

$$\lambda(t) = \frac{f(t)}{1-\sum_{s=1}^{t-1} f(s)} \tag{3.29}$$

The duration-type formulation (Equation 3.28) has both operational and meaningful advantages over the general model formulation because the termination rate function $\lambda(t)$ is frequently easier to parameterize, estimate, and interpret than the density function itself. These advantages are especially clear when modeling processes with duration-related conditionality. In addition, having the termination rate $\lambda(t)$ as an analytical function of t makes the duration model equally practical for any units of t.

Formulation of the duration model as a discrete choice model employs the following analytical form, assuming a multinomial logit model in this case:

$$P(t) = \frac{\exp(V_t)}{\sum_s \exp(V_s)} \tag{3.30}$$

where V_t denotes the utility function that is a linear-in-parameters function of independent variables:

$$V_t = \sum_k \beta_{kt} x_{kt} \tag{3.31}$$

where

$k \in K$ = household-, person-, zonal-, and duration-related variables;

x_{kt} = values of the variables for each alternative; and

β_{kt} = coefficients for the variables.

There is again a direct correspondence between the choice model (Equation 3.30) and the general-form density function (Equation 3.27). Any choice model has the corresponding density function calculated by Equation 3.30, and any density function (Equation 3.27) has an underlying set of utilities that are calculated by the following formula:

$$V_t = \ln f(t) \tag{3.32}$$

As in the case of duration models, discrete choice models (Equation 3.30) have advantages over the general formulation (Equation 3.27) because utility expressions (Equation 3.31) are easier to parameterize, estimate, and interpret than the density function itself. However, when the utility expression (Equation 3.31) is formulated in a general way with all alternative-specific coefficients and variables, the choice model (Equation 3.30) gets more complex with the addition of temporal resolution, which is not the case with the duration model (Equation 3.28). Also, the multinomial-logit formulation with independent alternative-specific variables suffers from the IIA property (independence from the irrelevant alternatives) with respect to those variables, ignoring the fact that the duration alternatives are naturally ordered.

Both of these deficiencies of the discrete choice formulation can be overcome using a certain specification of the utility function (Equation 3.31). This specification stems from an analogy that can easily be established between the duration model (Equation 3.28) and discrete choice model (Equation 3.30). Consider a ratio of densities for two subsequent points in time stemming from the two models, and restrict it to be equal in both cases:

$$\frac{P(t+1)}{P(t)} = \frac{\lambda(t+1)\times[1-\lambda(t)]}{\lambda(t)} = \exp(V_{t+1} - V_t) \tag{3.33}$$

Equation 3.33 contains several interesting and analytically convenient particular cases that lead to operational models that can be equally written and estimated in either duration form (Equation 3.28) or discrete choice form (Equation 3.30). Consider only one (actually, the simplest) case that corresponds to a duration model with a constant termination rate λ. With this assumption, Equation 3.33 is simplified to the following formula:

$$\exp(V_{t+1} - V_t) = 1 - \lambda \tag{3.34}$$

This means that there is a constant decrement in the utility function for each subsequent time point compared with the previous one, and it is the equivalent of the constant termination rate parameter of the duration model. The negative utility increment corresponds to the value of $1 - \lambda$ that is less than one. To ensure that the utility increment is independent of the time point, the variables x_{kt} and coefficients β_{kt} should be set in the utility expression (Equation 3.31) in a specific way. One of the possible ways to do this is to define all coefficients as

generic across duration alternatives ($\beta_{kt} = \beta_k$), while the variables are assumed to have the following form:

$$x_{kt} = t \times x_k \tag{3.35}$$

This formulation for the variables is not very restrictive since most of the household, person, and zonal characteristics in the TOD choice model are naturally generic across time alternatives. However, it is not true for network LOS variables that vary by TOD and should be specified as alternative specific. These variables, which are essentially time specific, violate the constant termination rate assumption. However, the discrete choice framework allows for easy hybridization of both types of variables (generic and time specific).

Using generic coefficients and variables of this type (Equation 3.35) creates a compact structure of the choice model in which the number of alternatives can be arbitrarily large (depending on the chosen time unit scale), but the number of coefficients to estimate is limited to the predetermined set *K*. These variables can be interpreted as continuous shift factors that parameterize the termination rate in such a way that a positive coefficient means the termination rate is getting lower, and the whole distribution is shifted to the longer durations. Negative values work in the opposite direction, collapsing the distribution toward shorter durations.

In the current research, the team also considered a nonlinear generalization of shift variables in the following forms:

$$x_{kt}^1 = t \times x_k; \quad x_{kt}^2 = t^2 \times x_k \tag{3.36}$$

where x_{kt}^1 and x_{kt}^2 are used in the utility function as independent variables with estimated coefficients β_k^1 and β_k^2 consequently. This extension of model structure allows for capturing some nonlinear effects, in particular saturation effects, in which the impact of a certain variable x_k is expressed in differential shifts along the duration time line. Essentially, the resultant multiplier for original variable x_k in the utility function ($t \times \beta_k^1 + t^2 \times \beta_k^2$) represents the timing profile for the impact of this variable.

In addition to nonlinear shifts in the current research, the team also applied various referencing and constraining schemes for shift variables. Referencing means that the shift is calculated relative to a certain point in time, and differential shifts can be applied for being earlier or later. Referencing can be formalized in the following way:

$$x_{kt}^3 = \min(t - t_k, 0) \times x_k; \quad x_{kt}^4 = \max(t - t_k, 0) \times x_k \tag{3.37}$$

where

t_k = reference time point (alternative) for the variable;
x_{kt}^3 = variable corresponding to shifts to an earlier time than the reference alternative; and
x_{kt}^4 = variable corresponding to shifts to a later time than the reference alternative.

Constrained shifts are only applied for a certain subset of adjacent alternatives, rather than for all 20 alternatives. For example, some peak-spreading effects can be localized within a specific peak period, such as 6:00–10:00 a.m., and are not relevant to the later hours.

In the process of model estimation, all types of shift variable transformations are applied in combination, including nonlinear effects, referencing, and constraining. The best combined form is defined by statistical fit and also by meaningful behavioral interpretations. The resulting impact of each variable x_k is referred to as its *timing profile.* It essentially singles out the impact of this variable on TOD choice. If the variable itself is a dummy (like female gender or income group indicator), the timing profile is expressed in utility units. This is the most common case with a straightforward interpretation. If the variable is continuous (like travel time or distance), then the interpretation of timing profile is more complicated and is expressed as a relative impact of each minute or mile of travel on TOD choice.

Time-of-Day Model Formulation for a Tour

Scheduling of an entire travel tour requires that the choice alternatives are formulated as tour departure-from-home (g) and arrival-at-home (h) hour combinations (g, h). Tour duration is derived as the difference between the arrival and departure hours ($h - g$). In the current research, tour duration incorporates both the activity duration and travel time to and from the main tour activity, including intermediate stops.

The tour TOD choice utility for a single tour can be operationalized in the following general form (Vovsha and Bradley 2004; Abou-Zeid et al. 2006; Popuri et al. 2008):

$$V_{gh} = V_g + V_h + D_{h-g} \tag{3.38}$$

where

g, h = departure-from-home and arrival-back-home times;
V_g = departure time choice–specific component;
V_h = arrival time choice–specific component; and
D_{h-g} = duration-specific component.

Departure hour– and arrival hour–specific components are estimated using generic shift-type variables (household, person, and zonal characteristics) according to Equations 3.35 and 3.36 with a limited set of TOD period-specific constants. Just as duration shift variables are multiplied by the duration of the alternative, departure shift variables are multiplied by the departure alternative, and arrival shift variables are multiplied by the arrival alternative.

Note that the index of the duration component is ($h - g$) rather than ($g \times h$), making the estimation procedure much simpler, since the number of duration alternatives is much less than the number of departure–arrival combinations. It should also be noted that none of the estimated components of the utility function (Equation 3.36) has an index with dimensionality ($g \times h$). Thus, the number of coefficients that

have to be estimated is in general fewer than the number of alternatives. This parsimonious structure, however, outperformed a model with a full set of ($g \times$ h) alternative specific constants (Vovsha and Bradley 2004).

Joint Time-of-Day and Mode Model Formulation

Model generalization to incorporate the mode choice dimension is straightforward and results in adding one more component to the utility function:

$$V_{ghm} = V_g + V_h + D_{h-g} + W_m(gh) \tag{3.39}$$

where

V_{ghm} = combined utility function for TOD and mode choice;

m = travel modes, car occupancy categories, and route types; and

$W_m(gh)$ = mode utility component with TOD-specific LOS variables.

Although the combined structure has a very large number of alternatives (210 × 13 = 2,730), the complexity of model estimation is approximately equal to the sum of efforts corresponding to TOD choice and mode choice due to the additive utility function. The mode-related component structures with all pertinent variables were adopted from the final version of estimated mode choice model discussed above. Thus, this mode choice construct is used as the starting point. All coefficients however, were reestimated in the more general choice framework, including travel time, cost, and reliability coefficients. It should be noted that all mode choice coefficients in the previously discussed mode choice model and this combined model were specified as generic across TOD periods to keep the model structure manageable. The mode-related utility components $W_m(gh)$ differ across TOD periods because the LOS variables were generated for each TOD (gh) specifically.

The mode choice coefficients were estimated simultaneously with the coefficients related to TOD choice. Despite the complexity of joint estimation, this method offers the significant advantage of the possibility of exploring different nested structures between the TOD and mode dimensions, as well as within each of them. Previously estimated tour TOD models of this type were fed by precalculated mode choice logsums for each TOD period. In the process of joint mode estimation, lower-level logsums were calculated automatically and adjusted according to the mode choice coefficient estimates.

Main Findings

This subsection summarizes the main model components and corresponding behavioral impacts that proved to be common for the New York and Seattle models. The main research question at this stage was whether adopting the more general framework of joint mode and TOD choice would change the main findings of the previous sections with respect to the seed form of the generalized cost of highway modes. The previously substantiated functional forms were subject to a series of additional statistical tests in which the utility function included both mode and TOD components.

The main conclusion that could be made at this stage was that for both the New York and Seattle models, the extension of the model to include TOD choice dimension in addition to mode dimension did not violate the main impacts of LOS and other variables. In particular, all main LOS components previously substantiated for more limited frameworks of route-type choice and mode choice proved to be statistically significant, with the right sign, and mostly with a similar magnitude in the more general choice context that included the TOD dimension. This confirms the main hypothesis of the C04 project, that there is a generic form of highway generalized cost that can combine mean travel time, cost, and travel time reliability (standard deviation of travel time per unit distance) measures, and that this form can be used as a seed component in the utility function through the entire hierarchy of main travel choices. This finding is encouraging, because using the same seed formulation for generalized cost from bottom up in the travel model system ensures consistency of the model system elasticities and responses to congestion and pricing.

In both the New York and Seattle models, the mode choice part of the utility for highway modes included trip-length effects on VOT through the interaction terms between travel time and distance substantiated previously for the route-type choice and mode choice frameworks. In this regard, all effects associated with trip distance captured in the mode choice framework discussed above (including nonlinear impacts on VOT) were preserved in the more holistic framework of integrated TOD and mode choices.

Several interesting direct effects of tour distance on TOD choice were captured with the New York data. The composite effect on departure time from home shows that with each additional mile of commuting, the probability of earlier departure will grow across all hours, with the strongest shift between the hours of 11 p.m. and 6 a.m. In a similar vein, each additional commuting mile proved to stretch the departure time for the return trip to home toward later hours, with the highest elasticity between 2 and 6 p.m.

The impact of car availability on the results of the New York and Seattle models proved to be very similar. The impact of household car availability on combined choice of TOD and mode was captured through the mode utility component, which proved to be very similar to the impact on pure mode choice. With the New York data, adding the TOD dimension did not change the main effects that were expressed through mode preferences by four car-sufficiency groups. In

the same vein, the mode preference effects included in the mode choice models based on the Seattle RP data were once again included in the mode and TOD choice models, and the results were much the same.

Income has several important impacts on joint choice of TOD and mode. The first effect relates to mode preferences. With the New York model, these impacts in the joint choice framework proved to be very similar to the mode preferences discussed above for the pure mode choice model. In the same vein, for the Seattle models, the team repeated the same tests that were done for the mode choice models reported above: segmenting the cost coefficient by income and vehicle occupancy versus assuming the same power function that was adopted for the analyses on the New York data. Similar to the tests with the New York model, the results were virtually unchanged from what was found for the mode choice–only models. The second income impact relates to schedule preferences. For example, the New York data showed that low-income commuters tend to have later schedules (departure from home after 9 a.m.) more frequently than medium- and high-income workers. The most prominent feature of high-income commuters is that they avoid very early starts (before 7 a.m.) compared with medium- and low-income workers.

Similar to the income variable, the joint travel variable has two intertwined effects on combined mode and TOD choice. The first effect is captured by the mode-specific utility components that enter each TOD choice alternative with the corresponding LOS variables. These effects are explained by the mode choice component, and they remain very similar in the joint mode and TOD choice formulation to the mode choice effects (specifically the subsplit between single-occupant vehicle [SOV] and high-occupancy vehicle [HOV]) discussed above. However, in the joint mode and TOD formulation, a second (direct) effect of car occupancy on TOD choice was captured with the New York model that is more related to carpool organization factors and the associated schedule constraints of the participants. Carpooling commuters are characterized by a later departure from home when very early hours, from 5 to 7 a.m., are avoided. In a similar vein, carpoolers avoid very late arrival times back home (after 7 p.m.), as late arrivals might not be convenient for at least some members of the travel party.

With respect to additional person characteristics, the team first summarized the impacts of person and household characteristics on the mode choice component, comparing it with the previously estimated pure mode choice formulations above. For the Seattle model, with cost divided by functions of income and vehicle occupancy, the team also included additive travel time variables specific to females and part-time workers, as well as an additive travel time variable multiplied by age minus 18 (set at a minimum of zero to apply only to those over 18 years of age). The results were somewhat less significant than were found in the mode choice–only models, with no significant differences in VOT related to gender, age, or part-time employment status. Second, there are significant mode preferences and TOD shift variables related to these characteristics, as reported in previous sections. Similarly, with the New York model, several TOD-related effects of person characteristics were captured in addition to the person variables included in the mode choice portion of the combined utilities. The most important distinction that strongly affects TOD choice for commuters is worker status. Part-time workers are characterized by later departure-from-home time than full-time workers. With respect to work tour duration, part-time workers are characterized by significantly shorter schedules than full-time workers. For arrival time back home, most of the effects will be derived as a composition of departure time and duration effects. For example, longer durations for full-time workers will naturally create later arrivals, all else being equal.

Travel time reliability was explored with respect to impacts on two travel dimensions. As discussed above, travel time unreliability, measured as a standard deviation of travel time per unit distance, affects choice of modes as it was found statistically significant in mode choice utilities for highway modes, in addition to the mean travel time. With the New York data, this effect becomes even more statistically significant when mode choice is considered jointly with TOD choice. In general, the greater the level of unreliability of highway travel time, the lower is the share of highway modes versus transit and other modes, all else being equal. It was also important to explore a possible direct impact of travel time reliability on TOD choice, in addition to the effect incorporated in mode choice logsums. For this purpose, with the New York model, travel time reliability in the measure was explored statistically as a shift variable in the TOD portion of the utility. The results confirmed two logical and statistically significant effects. The first effect relates to the shift of commuting departure time to hours earlier than 8 a.m., which is progressively stronger for each earlier hour. This statistical evidence fully confirms the fact that commuters have to take into account a certain extra (buffer) time in the presence of travel time unreliability. A similar symmetric effect was found for arrival time back home after work. Travel time unreliability resulted in later arrivals, with a progressive effect between 5 and 9 p.m.

Several effects associated with urban density and land use type were explored with the New York data. Some effects were already incorporated in the mode choice utilities as discussed above. The effects remained stable after extension of the choice dimensions to include TOD in addition to mode. The additional direct effect on TOD choice captured by the Manhattan dummy is associated with a significantly longer duration of work tours. Manhattan jobs are characterized primarily by office and managerial occupations that are associated with longer durations and more flexible arrangements like a compressed work week.

Basic Specification, Segmentation, and Associated Value of Time

The main conclusion that could be made at this stage was that for both the New York and Seattle models, the extension of the model to include a TOD choice dimension in addition to mode dimension did not violate the main impacts of LOS and other variables. In particular, all main LOS components previously substantiated for more limited frameworks of route-type choice and mode choice proved to be statistically significant, with the right sign, and mostly with a similar magnitude, in a more general choice context that included the TOD dimension. This finding confirms the main hypothesis of the C04 project that there is a generic form of a highway generalized cost function that can combine mean travel time, cost, and travel time reliability (standard deviation of travel time per unit distance) measures and that this form can be used as a seed component in the utility function through the entire hierarchy of main travel choices. This finding is encouraging because using the same seed formulation for generalized cost from bottom up in the travel model system ensures consistency of the model system elasticities and responses to congestion and pricing.

There are some particularly subtle effects associated with a joint consideration of mode and TOD choice compared with a pure mode choice model with fixed TOD. The primary difference, which manifests itself more strongly in a congested area like New York, is that when the TOD choice dimension becomes endogenous, it makes congestion-averse behavior more explicit. In the New York model it resulted in a stronger impact of travel time reliability. In general, due to a strong interdependence between mode and TOD choice, it is desirable to estimate these models jointly rather than sequentially. The current research has proven that this is both possible and practical, even though this practice results in a complicated choice structure, with a large number of alternatives, a large number of coefficients in the combined utility functions, and several nesting levels to explore. In the model application, the model can still be broken into a sequence of submodels by nesting levels that is equivalent to a fully joint model if the lower-level logsums are properly carried up.

Nonlinear Level of Service, Trip Length, and Location Effects

In both the New York and Seattle models, the mode choice part of the utility for highway modes included trip-length effects on VOT through the interaction terms between travel time and distance that were substantiated previously for the route-type choice and mode choice frameworks. In this regard, all effects associated with trip distance captured in the mode choice framework discussed above (including nonlinear impacts on VOT) were preserved in the more holistic framework of integrated TOD and mode choices. With the Seattle RP data, compared with the mode choice models results, the curve is less pronounced for HBO and reaches a maximum at a higher distance (around 40 miles) for HBW.

In addition, several interesting direct effects of tour distance on TOD choice were captured with the New York data. Commuting distance proved to have a direct impact on departure time from home and arrival time back home that was captured by linear and squared shift variables. The composite effect on departure time from home shows that with each additional mile of commuting, probability of earlier departure will grow across all hours, with the strongest shift between the hours of 11 p.m. and 6 a.m. In a similar vein, each additional commuting mile proved to stretch the departure time from home toward later hours, with the highest elasticity between 2 and 6 p.m.

Impact of Congestion Levels

In both the New York and Seattle regions the extension of the choice dimensions to TOD and mode did not significantly change the previous results with respect to auto time segmentation by congestion levels. Overall, the results were statistically unstable or insignificant, or both. Thus, these results only reinforced the decision to apply direct measures of travel time (un)reliability, like standard deviation of travel time per unit distance, rather than use indirect measures, like auto time weights differentiated by congestion levels.

Impact of Household Car Availability

The impact of car availability yielded very similar results with both the New York and Seattle models. The impact of household car availability on combined choice of TOD and mode was captured through the mode utility component in a way that proved to be very similar to the impact on pure mode choice. With the New York data, adding the TOD dimension did not change the main effects that were expressed through mode preferences by four car-sufficiency groups. In the same vein, the mode preference effects included in the mode choice models based on the Seattle RP data were once again included in the mode and TOD choice models, and the results were much the same, with the exception that many of the effects were estimated even more significantly.

Impact of Household or Person Income

Income has several important impacts on joint choice of TOD and mode. The first set of impacts relates to mode preferences. With the New York model, these impacts in the joint choice framework proved to be very similar to the mode preferences discussed above for the pure mode choice model. In the same

vein, for the Seattle models, the same tests were repeated that were done for the mode choice models that are reported above: segmenting the cost coefficient by income and vehicle occupancy as opposed to assuming the same power function that was adopted for the analyses on the New York data. Similar to the tests with the New York model, the results were virtually unchanged from what was found for the mode choice–only models.

The second set of income impacts tested relates to schedule preferences. For example, the New York data showed that low-income commuters tend to have later schedules (departure from home after 9 a.m.) more frequently than medium- and high-income workers. The most prominent feature of high-income commuters is that they avoid very early starts (before 7 a.m.) compared with medium- and low-income workers. This finding correlates with the nature of corresponding occupations and schedule flexibility.

Impact of Joint Travel

Similar to the income variable, the joint travel variable has two intertwined effects on combined mode and TOD choice. The first effect is captured by the mode-specific utility components that enter each TOD choice alternative with the corresponding LOS variables. These effects are explained by the mode choice component and, for the New York model, they remain very similar in the joint mode and TOD choice formulation to the mode choice effects (specifically, the subsplit between SOV and HOV), as discussed above. In a similar way, with the Seattle model, the carpooling impacts were virtually unchanged from what was found for the mode choice–only models discussed above, with less sensitivity of the cost coefficient to vehicle occupancy than what is represented in the assumed power functions, particularly in the HBO models.

However, in the joint mode and TOD formulation, a second (direct) effect of car occupancy on TOD choice was captured with the New York model. Although the first (logsum-related) effect is sensitive to LOS variables and corresponding policies like HOV and high-occupancy toll (HOT) lanes, the second effect is more related to carpool organization factors and the associated schedule constraints of the participants. Carpooling commuters are characterized by a later departure from home when very early hours, from 5 to 7 a.m., are avoided. In a similar vein, carpoolers avoid very late arrivals back home (after 7 p.m.), because late arrivals might not be convenient for at least some members of the travel party.

Impact of Gender, Age, and Other Person Characteristics

This section first summarizes the impacts of person and household characteristics on the mode choice component, comparing it with the previously estimated pure mode choice formulations discussed above. For the Seattle model, with cost divided by functions of income and vehicle occupancy, the team also included additive travel time variables specific to females and part-time workers, as well as an additive travel time variable multiplied by age minus 18 (set at a minimum of zero to apply only to those over 18 years of age). The results were somewhat less significant than those found in the mode choice–only models, with no significant differences in VOT related to gender, age, or part-time employment status.

Second, there are significant mode preference and TOD shift variables related to these characteristics, as reported in previous sections. Similarly, with the New York model, several TOD-related effects of person characteristics were captured in addition to the person variables included in the mode choice portion of the combined utilities. The most important distinction that strongly affects TOD choice for commuters is worker status, for which three main person types (full-time worker, part-time worker, and nonworker) are considered. The last category includes some commuters for job interviews, occasional work, and volunteers. Part-time workers and nonworkers are characterized by later departure-from-home time than full-time workers. With respect to work tour duration, both part-time workers and nonworkers are characterized by significantly shorter schedules than full-time workers. For arrival time back home, most of the effects will be derived as a composition of departure time and duration effects. For example, longer durations for full-time workers will naturally create later arrivals, all else being equal. However, in addition to the derived effects for full-time workers, one direct arrival time–related effect proved to be significant. Full-time workers rarely arrive back home before 3 p.m., in contrast to part-time workers and nonworkers.

Incorporation of Travel Time Reliability and Value of Reliability Estimation

Travel time reliability was explored with respect to impacts on two travel dimensions. As discussed above, travel time unreliability, measured as a standard deviation of travel time per unit distance, affects choice of modes as it was found statistically significant in mode choice utilities for highway modes, along with the mean travel time. With the New York data, this effect becomes even more statistically significant when mode choice is considered jointly with TOD choice. In general, the greater the level of unreliability of highway travel time, the lower is the share of highway modes versus transit and other modes, all else being equal. Through TOD-specific mode choice logsums, this impact also has an effect on TOD choice. However, the results for the Seattle model were less successful. For the HBW models, the reliability variables all have the incorrect sign, except when included as the buffer travel time (90th percentile minus

median) divided by distance, which has a significant negative coefficient. For the HBO models, the buffer travel time variable is again the only reliability variable with a significant negative coefficient, but this time when *not* divided by trip distance. It may be concluded that the distribution of level of congestion and associated variation in travel time reliability measures in the Seattle data was not rich enough.

It was also important to explore a possible direct impact of travel time reliability on TOD choice in addition to the effect incorporated in mode choice logsums. For this purpose, with the New York model, the travel time reliability measure was explored statistically as a shift variable in the TOD choice utility (departure from home and arrival back home components), in addition to inclusion of travel time reliability in the mode choice logsum. The results confirmed two logical and statistically significant effects. The first effect relates to the shift of commuting departure time to hours earlier than 8 a.m., which is progressively stronger for each earlier hour. This statistical evidence fully confirms the fact that commuters take into account a certain extra (buffer) time in the presence of travel time unreliability. A similar symmetric effect was found for arrival time back home after work. Travel time unreliability resulted in later arrivals, with a progressive effect between 5 and 9 p.m.

Impact of Urban Density and Land Use

Several effects associated with urban density and land use type were explored with the New York data. Some of the effects were already incorporated in the mode choice utilities as discussed above. The effects remained stable after extension of the choice dimensions to include TOD in addition to mode.

The most prominent new effect on TOD choice was associated with a simple Manhattan job dummy that captured the principal difference between commuting to Manhattan and the rest of the metropolitan area. The additional direct effect on TOD choice captured by the Manhattan dummy is associated with a significantly longer duration of work tours. This spans durations from very short to 14 hours. Two main behavioral mechanisms can explain this phenomenon. First, Manhattan jobs are characterized primarily by office and managerial occupations that are associated with longer durations and more flexible arrangements like a compressed work week. Second, in this analysis the team operated with the entire tour duration (from departure from home until arrival back home), rather than with duration of the work activity itself. Thus, additional activities (stops) on the way to and from work come into play. Logically, commuting tours to Manhattan are characterized by a higher frequency of stops, primarily in Manhattan, because of the great variety of opportunities for shopping and discretionary activities there.

Route Type, Time-of-Day, and Mode Choice: Stated Preference Framework

This section summarizes the findings from models estimated on data from the Seattle Puget Sound Regional Council (PSRC) mode choice SP experiment, the San Francisco County Transportation Authority (SFCTA) cordon pricing study SP, and the Los Angeles HOT lane SP study. These three SP experiments are described in Chapter 2, and Table 3.8 summarizes their main design characteristics. Some key differences between the experiments are as follows:

- The Seattle and Los Angeles experiments recruited people who had actually made recent trips in relevant highway corridors in the region, and then presented experiments (sent via a survey form customized to their actual trip) that offered hypothetical tolled options in that corridor. For the Los Angeles experiment, the tolled option was offered as a HOT lane or express lane alongside free general-purpose

Table 3.8. Summary of Design Characteristics of Three SP Experiments

Characteristic	Seattle SP	Los Angeles SP	San Francisco SP
SP choice context	Introduction of tolls on route (general)	Introduction of HOT or express lanes	Introduction of toll to enter downtown area
Recruitment method	Recent trips on relevant highways, from HH travel survey sample	Recent trips on relevant highways, from telephone recruit survey	Recent trips to downtown SF, recruited at parking locations
Offered nontolled auto alternative?	Yes, could be on a different highway	Yes, free general lanes on same highway	No, all auto trips to downtown pay toll
Offered different prices in the off-peak period?	Yes, same peak periods for all respondents	Yes, varied peak toll periods across sample	Yes, varied peak toll periods across sample
Offered transit mode alternative?	No	Yes	Yes
Included a travel time reliability variable?	Yes, varied frequency of extra delay, fixed at 15+ minutes duration	No	Yes, varied duration of delay, presented as “1 in 5” or “1 in 10” trips

Note: HH = household.

lanes. In the Seattle experiment, the free option could be on a different route, requiring a different travel distance;

- The San Francisco experiment recruited people who made recent auto trips and parked downtown, and then presented hypothetical options with a cordon toll charged to enter the downtown area. The experiment was customized to the actual trips and presented via computer-based interview screens, either in-person on laptops or via the internet. No free auto alternative was offered;
- All three experiments offered different prices in the peak and off-peak periods, but the San Francisco and Los Angeles experiments also customized and randomly varied the definition of the peak period across the sample in order to better estimate TOD switching preferences;
- The San Francisco and Los Angeles experiments offered a transit alternative, but no transit option was included in the Seattle experiment; and
- The Seattle and San Francisco experiments included a travel time variability and reliability attribute, but the Los Angeles experiment (by design) did not. For the Seattle SP, the duration of extra delay was fixed at 15 minutes and the frequency was varied. For the San Francisco SP, the frequency of extra delay was randomly set at either one in five trips or one in 10 trips for each respondent, and the duration of the extra delay was varied within respondents.

Basic Specification, Segmentation, and Associated Value of Time

The team attempted to make the framework for estimation of the SP data sets somewhat consistent with the framework used in the preceding chapters for the RP-based analyses. The team started with basic models and then incrementally added detail about specific SP attributes and segmentation variables. In contrast to RP data models, however, the models estimated for SP data are largely determined by the design of the SP experiment; that is, one can include only those choice alternatives and LOS attributes that were portrayed in the choice scenarios. For example, a variable for travel time reliability and variability can be included only if that attribute was explicitly included in the SP design. Furthermore, all of the variables that were included in the SP experimental design must be included; otherwise, their omission may bias the estimates of the included attributes.

Table 3.9 summarizes results for the Seattle, San Francisco, and Los Angeles SP experiments for a basic model specification including only the basic SP attributes, along with appropriate alternative-specific constants and nesting logsum parameters. For each experiment, two models are shown: one for work trips and another for nonwork trips. In this model specification, there is no segmentation by income or auto occupancy, and all relationships are assumed to be linear. For purposes of comparison, Table 3.9 generally shows only results in the form of ratios of the coefficients. Unless otherwise stated, the coefficient estimates were significantly different from zero (although the team did not account for repeated measurements within respondents in estimation, so the standard errors will be somewhat underestimated). The values of all estimated coefficients and *t*-statistics can be found in the appendix.

All SP models were estimated as nested models. For the experiments that offered both tolled and nontolled auto alternatives (Seattle and Los Angeles), the tolled and nontolled route options were nested under each TOD period, and the logsum parameters were all significantly less than 1.0, generally around 0.4. Although one would not necessarily expect to estimate the same nesting coefficients for RP and SP data (due to differences in the way the choice sets are specified), it is interesting that this same nesting of route type under time period was also found for the New York RP and Seattle Traffic Choices data sets.

For the experiments that offered a transit alternative (San Francisco and Los Angeles), it was also found best to nest the auto alternatives across time periods and put the auto versus transit choice at the highest level. The logsum parameters were estimated at around 0.6 for the HBW models, but for HBO, there was some instability in estimating logsum coefficients; coefficients constrained to 0.5 gave a better fit than a nonnested model. Note that this nesting of TOD under mode is different from the results obtained for the Seattle and New York RP data. The SP experiments, however, were offered only to auto users in the context of actual trips they had made by auto, so there will naturally be less tendency to choose transit than one would find in a representative RP sample.

A key finding of the SP experiments is the overall willingness to pay for auto travel time savings in the form of the ratio of the auto in-vehicle and cost coefficients. Table 3.9 shows very similar estimates for the Seattle and Los Angeles data sets; both are in the range of \$11–\$12/hour for HBW trips and \$9–\$10/hour for HBO trips. These values are in a range typically estimated for highway users in the context of a toll project. For the San Francisco SP, however, the team estimated values that are about 50% higher than for the other experiments (in the range of \$15–\$18/hour). This result could be due to higher incomes, on average, for those who travel to downtown San Francisco; this possibility is explicitly tested in the next section. It could also be due, however, to the different context of cordon pricing. There is no nontolled option except for switching mode (or destination), so auto users may be more willing to pay a toll, particularly in the case of downtown San Francisco, where the cost of parking is already high by comparison.

It is worth noting that all three SP experiments show very similar overall VOT for those making work trips versus nonwork trips, with VOT for work trips 10%–20% higher in each case. This is in contrast to the team's findings for the two RP studies: Although the New York RP study found higher

Table 3.9. Basic Specifications for Three SP Experiments

	SP Experiment					
	Seattle	LA	SF	Seattle	LA	SF
	HBW	HBW	HBW	HBO	HBO	HBO
Summary of Results	**psspw1**	**laspw1**	**sfspw1**	**psspn1**	**laspn1**	**sfspn1**
VOT, auto in-vehicle time ($/hour)	12.0	11.2	17.7	9.9	9.4	15.8
Values in Equivalent Minutes Auto In-vehicle Time						
Toll route constant (min)	7.0	13.5	NA	8.8	14.5	NA
Distance to avoid toll (min/mi)	0.47	NA	NA	1.02	NA	NA
Average extra delay (min/min)	2.42	NA	0.42	2.65	NA	2.91
Shift earlier in a.m. (min/min)	0.17	0.40	0.21	0.05	0.62	0.29
Shift later in a.m. (min/min)	0.28	0.91	0.66	0.25	0.97	0.09
Shift earlier in p.m. (min/min)	0.20	0.36	0.09	0.03	0.39	0.27
Shift later in p.m. (min/min)	0.10	0.79	1.71	0.00	0.72	0.24
Transit total travel time (min/min)	NA	1.36	NA	NA	0.99	NA
Transit in-vehicle time (min/min)	NA	NA	0.72	NA	NA	1.10
Transit out-of-vehicle time (min/min)	NA	NA	1.09	NA	NA	0.97
Transit service frequency (min/min)	NA	0.43	NA	NA	0.73	NA
Transit transfers (min/transfer)	NA	15.7	9.7	NA	15.9	15.9
Transit mode constant (min)	NA	57.4	28.0	NA	193.1	35.6
Nesting Logsum Parameters						
Toll or nontoll nested under TOD	0.402 (–4.4)	0.387 (–15.2)	na	0.463 (–6.2)	0.264 (–17.8)	na
OTOD (*t*-statistic versus 1.0)						
TOD periods nested under modes (*t*-statistic versus 1.0)	na	0.581 (–3.5)	0.669 (–3.8)	na	0.50 (constr)	0.50 (constr)
Summary Statistics						
Observations	1,355	2,976	2,357	1,507	2,932	2,722
Rho-squared with respect to 0	0.247	0.297	0.166	0.247	0.276	0.109
Final log likelihood	–1,414.4	–3,907.5	–2,723.7	–1,574.0	–3,961.5	–3,360.4

Note: LA = Los Angeles; SF = San Francisco; HBW = home-based work trips; HBO = home-based other trips; na = not applicable; and NA = not available.

VOT for work trips, the Seattle RP study found higher VOT for nonwork trips. Standard practice is to use much higher VOT for work trips than for nonwork trips, but such a result is rarely found in SP-based studies. According to household welfare economics, one would expect VOT for any personal trip, commuting or otherwise, to be proportional to the value of spending time in the leisure activity that the saved time would be devoted to (i.e., the value of leisure time at the margin) relative to the value of spending time driving a vehicle. This, however, assumes that travelers can schedule their travel and activities predictably, and that time is fully substitutable between activities. In reality, there are a number of reasons why those conditions may not always hold:

- There may be unexpected delays or conditions that cause time to be taken from more valuable activities, such as work or more highly valued leisure activities;
- In response to unreliability, travelers may leave more of a buffer time between activities, particularly those activities with a high penalty or disutility for arriving late. Adding buffer time to travel times results in a suboptimal scheduling of activities and, possibly, a lower-valued use of time savings; and
- All alternative uses of time are not available at all times of day. For example, many leisure activities may not be possible during the early morning hours, and it may not be possible to shift work schedules to allow travel time saved in the morning to be used later in the day. Leisure activities often need to be scheduled in coordination with others, which further limits the possibility to schedule them at any given TOD.

These various conditions may differ for work travel relative to nonwork travel. In particular, the effects of reliability and

unexpected delays will, on average, tend to be stronger for work trips than for leisure trips (although they may also be quite strong for specific types of nonwork trips). This means that the more that reliability can be explicitly accounted for in models, the more one would expect to see that the net value of travel time is similar between work and nonwork travel.

Toll Route Constant

In Table 3.9, the remaining estimates are reported as ratios relative to the auto in-vehicle time coefficient, normalized to equivalent minutes of travel time. The first row is for the toll route constant, which was significant and negative for both travel purposes in the Seattle and Los Angeles experiments. (All auto alternatives were priced in the San Francisco case, so no constant was estimated.) Negative toll route constants are typically found in SP studies, and often in RP studies, as well (including the RP studies on the New York and Seattle data described above). All else equal, the constant on the tolled route is equivalent to about 8 minutes of extra in-vehicle time for the Seattle SP, and about 14 minutes in the Los Angeles SP.

Detour Distance to Avoid Tolls

In the Seattle SP, the nontolled path was not always on the same facility and could involve driving additional distance. Each mile of extra distance was valued negatively, above and beyond the time required to drive it. The value is twice as high for nonwork as for work trips, suggesting that work trips are more willing to search for alternative paths to avoid tolls (perhaps due to more familiarity with alternative routes). The value for work trips is equivalent to about 0.5 minutes per mile. If the average speed on the extra distance were 30 mph, then it would take 2 minutes to drive each extra mile, so this extra term would increase the disutility by 25% in that case.

Extra Delay and Reliability

As mentioned above, the Seattle and San Francisco SP experiments both included attributes related to the frequency and duration of extra delays above the usual travel time. Ideally, respondents are presented with a distribution of day-to-day travel times to obtain estimates comparable to the RP-based results given in the previous sections. Some recent SP studies have done this by presenting respondents with a series of five or 10 possible travel times for each alternative instead of a single time. Both the Seattle and San Francisco experiments opted for simpler approaches. The Seattle SP obtained more significant estimates, suggesting that people find it easier to understand the approach with fixed duration of extra delay and comparing different frequencies (e.g., one in 10 trips versus one in 20 trips), rather than vice versa. In either case, it is possible to multiply the frequency and duration to obtain the expected minutes of extra delay, which is somewhat comparable to a standard deviation measure (if travel times were never shorter than the typical time). Apart from the work trip result for the San Francisco SP, which was not statistically significant, the other three estimates indicate that each expected minute of extra delay is equivalent to about 2 to 3 minutes of expected travel time. If the expected extra delay is comparable to the standard deviation, this result suggests a reliability ratio between two and three, which does not seem out of the question. In general, it is expected that an extra delay minute would be valued more than a minute of the mean travel time (Li et al. 2010; Concas and Kolpakov 2009). Standard deviation is a symmetric measure that reflects both cases of being early and late. The relationship between standard deviation and expected lateness depends on the distribution of travel time. For symmetric distributions around the mean, the expected lateness is equal to $1/\sqrt{2}$ of the standard deviation.

Shifting Out of the Peak Pricing Period

In addition to the various period-specific constants that are reported in Appendix A, all of the SP models include shift variables for people who actually traveled in the peak pricing period and who could shift out of it to pay a lower toll. In contrast to RP-based TOD shift variables that are cross-sectional in nature, the SP-based variables are pseudolongitudinal, measuring before and after responses to hypothetical system changes. In general, the Seattle measures are lower in magnitude and less statistically significant than measures from the other studies, primarily because only one definition of the peak pricing period was used for the entire sample in the Seattle experiment. This restriction does not provide much variation for identifying shifting preferences; in contrast, the other experiments used peak period definitions that were semirandomly customized to respondents' trips.

In general, the largest resistance to shifting trip times is to shift later in the a.m. peak, particularly for work trips. As many individuals have to be at work by a specific time, this result makes sense. For the Los Angeles and San Francisco experiments, the disutility of each minute of shifting later in the a.m. is almost as large as the disutility of a minute of travel time, which is quite a high result. In the p.m., there seems to be somewhat more resistance to shifting later than shifting earlier. These results will also tend to vary depending on the current departure time; for example, someone already traveling at 7 a.m. may be less willing to shift earlier than someone traveling at 8 a.m. They may also be nonlinear, with some travelers having thresholds at which the shift becomes more difficult to schedule. These aspects of behavior are investigated in Chapter 4.

Shifting to Transit

A number of different variables were used to represent the transit alternative, which is necessary to give respondents some clear idea of how attractive the transit alternatives would be for their particular trips. This includes travel time broken down into components, as well as transfers, frequency, and fare. Because the SP samples only included actual auto users and transit was offered as an alternative to paying tolls, it would not be expected that these experiments would provide the most accurate or representative measures of the value of various transit service levels; RP and SP data from representative samples are better for that. Nevertheless, it can be seen that each minute of travel by transit has estimated values fairly similar to the value of auto time. Each transit transfer has a disutility equal to about 15 minutes of extra travel time, which is in a range typically estimated. The most interesting result is perhaps for the residual mode constant for transit. As one might expect, the resistance to switching to transit, all else equal, is higher for the Los Angeles experiment than for the downtown San Francisco experiment, equivalent to about 60 minutes and 30 minutes of auto travel time, respectively, for work trips. There is a wide selection of transit options into downtown San Francisco, whereas many of the highway corridors studied in the Los Angeles experiment currently have very little transit service, and many of the Los Angeles respondents may have never used transit for those trips. The resistance to shifting to transit is particularly high for the Los Angeles nonwork trips.

Impact of Household or Person Income

The discussion in this section and the following sections reflects the estimation of a second set of models that included the following additional variables:

- Segmentation of the travel cost coefficient by income quartile;
- Segmentation of the auto in-vehicle time coefficient by occupancy (SOV versus HOV); and
- Segmentation of the TOD shift variables by actual TOD, plus estimation of nonlinear functions rather than simple linear effects.

The income effects on the cost coefficient and VOT are shown in detail in the appendix and are summarized in Table 3.10 and Figure 3.6. The differences between income groups are generally significant and in the expected direction, with cost having a more negative coefficient (lower VOT ratio) for lower-income groups. When plotted in the graph, all of the experiments and purposes show a fairly similar trend of increasing VOT across quartiles. (Roughly, the lowest quartile is below $30,000, the second is $30,000–$60,000, the third is $60,000–$100,000, and the highest is over $100,000; these division vary somewhat by sample.) Curiously, after the change of model specification, the San Francisco sample shows somewhat lower VOT than the other regions, except in the highest income quartile. The graph also uses dotted lines to show approximately what the curve would look like if the VOT trend conformed to a power function of 0.4, 0.5, 0.6, or 0.7. In general, the curves of 0.4 and 0.5 seem closest in slope to the estimated trends, slightly lower than the exponents of 0.5 and 0.6 that were assumed in the RP analyses for this project.

Impact of Joint Travel

After various specification tests, in the SP experiments the effect of vehicle occupancy (SOV versus HOV) on willingness to pay was captured better by segmenting the travel time coefficient rather than the cost coefficient. As summarized in Table 3.11 (the detailed estimation results are given in Appendix A), willingness to pay is higher for HOV than for SOV in all models, particularly for nonwork trips. In general, however, the effect is less than linear with vehicle occupancy, and less than the power function exponents assumed in the

Table 3.10. Cost Coefficients and VOT by Income in Three SP Experiments

	SP Experiment					
	Seattle	LA	SF	Seattle	LA	SF
	HBW	HBW	HBW	HBO	HBO	HBO
Income Quartile	psspw2	laspw2	sfspw2	psspn2	laspn2	sfspn2
Lowest	8.6	7.6	6.0	6.7	6.9	4.4
Second	10.2	8.3	7.5	7.9	7.2	6.7
Third	13.2	10.5	9.2	9.1	8.5	7.1
Highest	17.7	16.2	11.6	12.7	10.3	16.6

Note: VOT is for auto in-vehicle time.

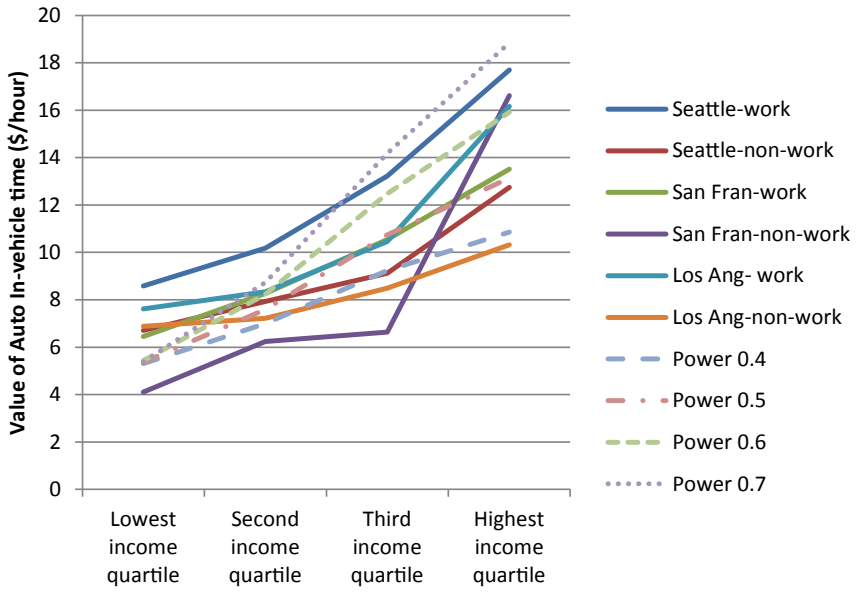

Figure 3.6. Summary of estimated VOT by income group.

RP analysis. Typically, SP samples only include the vehicle driver and not the other vehicle occupants, and it is not always clear to what extent respondents are answering only on their own behalf and to what extent they are answering for the entire traveling party, particularly with regard to sharing payment of tolls or other travel costs. As a result, SP results may be more accurate and representative for SOV trips than for HOV trips.

Incorporation of Departure Time Shift Effects

The San Francisco and Los Angeles SP data sets allowed detailed analysis on the willingness to shift out of peak pricing periods as a function of the toll level, the amount of time shift necessary, and the current time of travel. The exact coefficients are given in the appendix, and the functions are plotted below. As an example, the two graphs in Figure 3.7 show the disutility of shifting departure time either earlier or later to avoid the a.m. peak pricing period for trips to work. The results indicate a stronger resistance to moving work departure time later versus moving it earlier, at least for smaller shifts. In the range of 0 to 45 minutes, the second chart has a slope of more than 1 minute versus SOV travel time; that is, each minute of moving departure time earlier has a disutility worth more than 1 minute of in-vehicle time. At higher levels, however, the slope flattens out. Presumably, once one is very late for work, additional shifts do not make as much difference.

The shifts in the curves for different actual departure times indicate that those who actually go to work very early in the morning are more resistant to changing departure time in either direction, earlier or later. It is understandable that these people would be more averse to shifting earlier, since they would need to start their day very early to do so. The fact that early risers are also more averse to moving later may be due to the fact that they have less flexible work schedules. That the same trend is found for both experiments provides evidence that this finding is not an anomaly.

For nonwork trips in the a.m. peak, the picture is not as clear, as shown in Figure 3.8. For the San Francisco experiment, more resistance is again found to moving later than moving earlier, but with a different picture by TOD. In this case, it is those who are already traveling later in the a.m. who are most resistant to moving even later. For moving earlier, there is very little difference related to the actual TOD. For Los Angeles, the pattern for nonwork trips in the a.m. looks more similar to the pattern for work trips. The resistance to

Table 3.11. Time Coefficients and VOT Segmented by Occupancy for Three SP Experiments

	SP Experiment					
	Seattle	LA	SF	Seattle	LA	SF
	HBW	HBW	HBW	HBO	HBO	HBO
Summary of Results	psspw2	laspw2	sfspw2	psspn2	laspn2	sfspn2
Ratio of shared ride (HOV) VOT	1.03	1.28	1.12	1.39	1.34	1.90

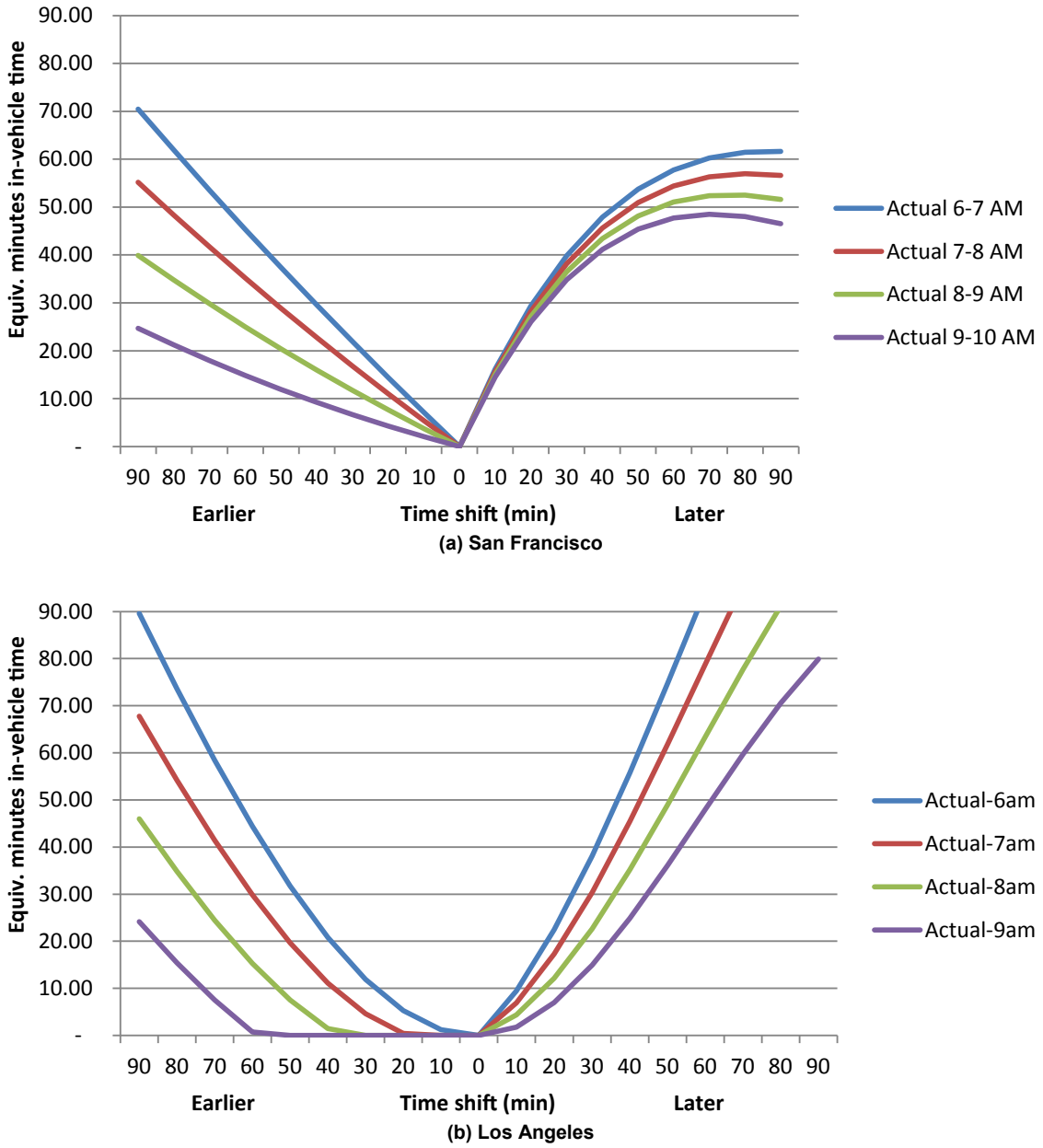

Figure 3.7. Resistance to shifting time to respond to TOD for a.m. peak work trips.

moving later levels off after about 45 minutes (the bend in the curve is an artifact of the cubic function adopted).

For nonwork trips in the p.m., both San Francisco and Los Angeles show similar sensitivities, with generally somewhat less resistance to shifting times than in the a.m. peak, as shown in Figure 3.9. For San Francisco, it is once again seen that those who travel earlier in the day are somewhat more resistant to changing times. The trips made earlier in the afternoon may be more likely to be for fixed appointments than for trips made after usual office hours. For Los Angeles, it can be seen that those traveling earlier are more averse to shifting earlier, but those traveling later in the afternoon are more averse to shifting later.

Incorporating Unobserved Heterogeneity

To investigate unobserved heterogeneity in route-type choice and TOD, data from the Seattle SP toll choice experiment were used. Four alternatives were available to individuals:

- Peak period + free;
- Peak period + toll;
- Nonpeak period + free; and
- Nonpeak period + toll.

The peak period occurred 6–9 a.m. or 3–7 p.m. Due to overlapping of the alternatives, the nesting structure shown in Figure 3.10 was assumed for modeling.

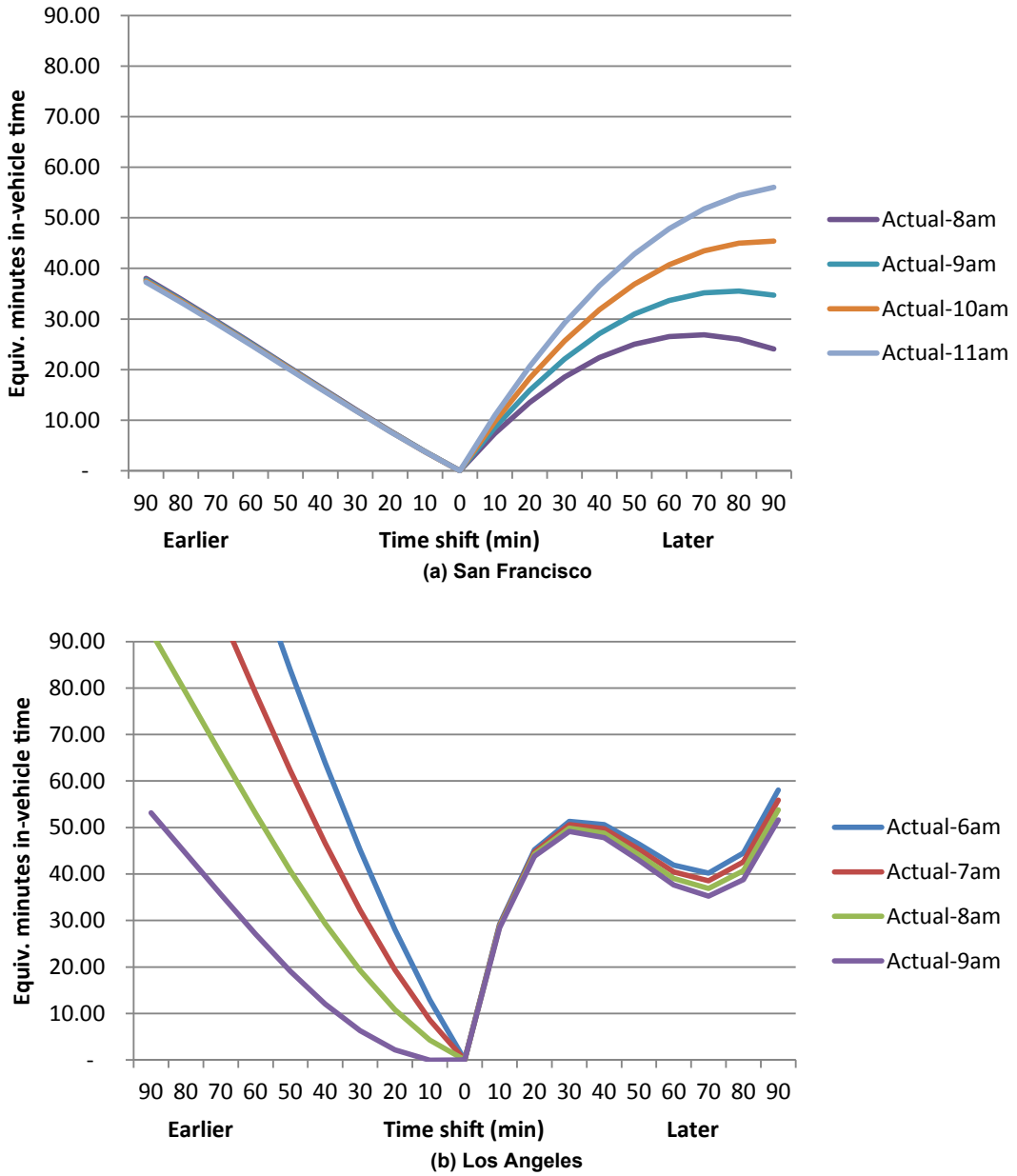

Figure 3.8. Resistance to shifting time to respond to TOD for a.m. peak nonwork trips.

The estimation results, which are shown in Table 3.12, illustrate some important insights that can be gained by accounting for unobserved heterogeneity in travel time response. By capturing the distribution of individuals' VOT, the proportion of the population with a specific VOT can be determined. In contrast with assuming all individuals have the same VOT represented by the average, Figure 3.11 shows that individuals vary greatly in their VOT.

By examining the estimation results in Table 3.12, the impact of accounting for unobserved heterogeneity in choice models is realized. First, notice that as more unobserved heterogeneity is captured, the log likelihood in general decreases. For a mixed logit model that captures serial correlation in addition to unobserved heterogeneity in the travel time coefficient, the log likelihood improves from –3095.5678 to –2789.5533. The log likelihood did not improve when just adding a random coefficient for travel time; this lack of improvement may be attributed to fixing the nesting parameter. Second, depending on the type of correlation and heterogeneity captured (e.g., randomness in the travel time coefficient or both random coefficient and serial correlation across observations), VOT varies over the population differently. Initially, by only capturing variation or heterogeneity in the travel time coefficient, the variance of VOT is larger, relative to whether both serial correlation and random travel time perception are captured. One possible explanation for

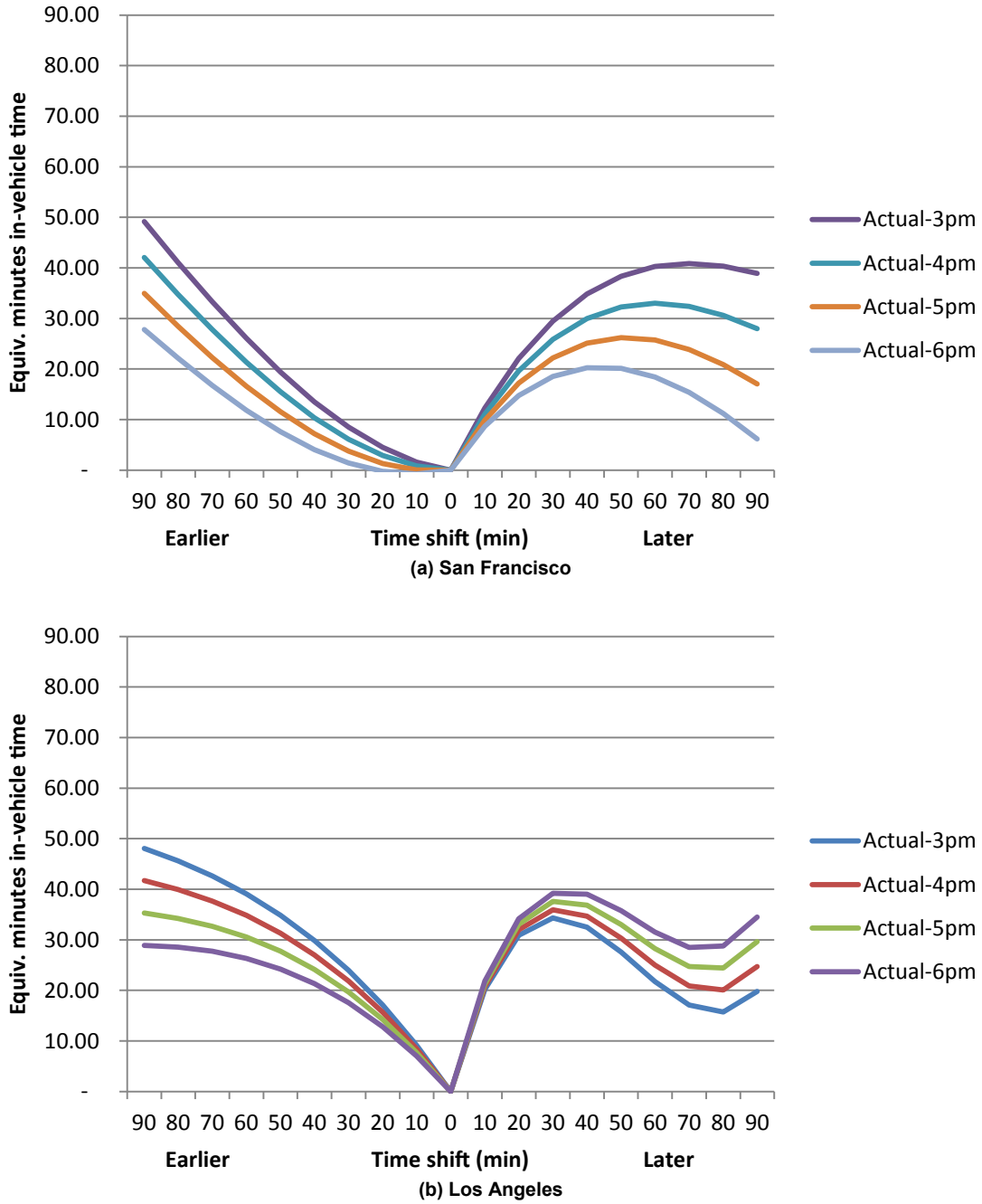

Figure 3.9. Resistance to shifting time to respond to TOD for p.m. peak nonwork trips.

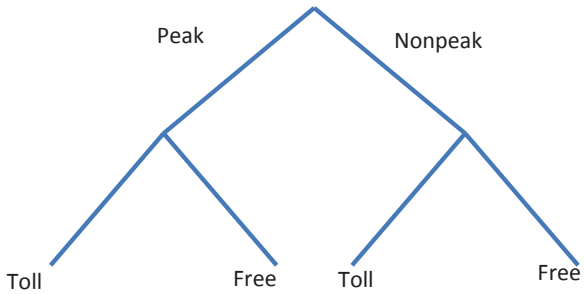

Figure 3.10. Nesting structure for Seattle SP TOD + route-type choice.

this larger variance is that more of the variance is captured by serial correlation. By not accounting for this serial correlation, VOT exhibits greater variance.

Examining the cumulative distribution functions of VOT under different assumed correlations shows that assuming only a random coefficient for travel time gives a steeper initial cumulative distribution relative to the case with both serial correlation and random coefficients.

An interesting result is that similar to the mode choice model, as more correlations are captured in the model, the variance of VOT decreases relative to the case in which no correlations are

Table 3.12. Estimation Results for RP Route-Type Choice

Model	Nested Logit		Mixed Logit		Mixed Logit	
Distribution	na		Lognormal		Lognormal	
Observations	2862		2862		2862	
Final Log Likelihood	–3041.6584		–3095.5678		–2789.55	
Rho-squared (const)	0.234		0.234		0.234	
Rho-squared (zero)	0.214		0.214		0.214	
Variable	**Coefficient**	**T-statistic**	**Coefficient**	**T-statistic**	**Coefficient**	**T-statistic**
Toll Cost ($)	–0.5580	–7.91	–0.5766	–6.54	–1.0110	–8.59
Toll Cost/Income ($/K$)	–6.9949	–2.65	–9.6423	–2.78	–9.6595	–2.10
Toll Cost∗#Passengers ($)	0.0660	2.03	0.0219	0.47	–0.2077	–3.29
Travel Time (min)	–0.1252	–13.36	–0.1821	na	–0.3558	na
Travel Distance (miles)	–0.0928	–3.53	–0.0887	–3.08	0.0263	0.55
Fraction of Times Late	–8.2644	–5.55	–8.8527	–10.16	–10.4306	–7.94
Fraction of Times Late Squared	6.564	2.52	7.0237	4.58	6.2875	2.32
Off-Peak∗actual minutes after 6 AM	–0.0152	–5.22	–0.0157	–5.21	–0.0202	–5.45
Off-Peak∗actual minutes before 9 AM	–0.0363	–12.08	–0.0372	–10.90	–0.0407	–10.35
Off-Peak∗actual minutes after 3 PM	–0.0083	–4.83	–0.0084	–4.79	–0.0106	–4.47
Off-Peak∗actual minutes before 7 PM	–0.0056	–3.34	–0.0060	–3.46	–0.0051	–2.11
Off-Peak∗actual off-peak	3.1297	43.41	3.1482	17.00	3.3892	13.99
Toll route constant	–1.0408	–11.85	–1.0755	–10.24	–1.4469	–8.76
Toll Nesting Parameter	0.3862	19.65	0.3862	na	0.3862	na
Error Term Parameters						
Variance of Beta-Travel Time	na	na	0.0835	na	0.0792	na
Variance Alternative 1	na	na	na	na	2.7751	1.98
Variance Alternative 2	na	na	na	na	3.7015	5.82
Variance Alternative 3	na	na	na	na	9.0000	na
Variance Alternative 4	na	na	na	na	2.2841	94.16
Covariance Alternative 1 Time Lag	na	na	na	na	1.1139	1.51
Covariance Alternative 2 Time Lag	na	na	na	na	1.9792	1.15
Covariance Alternative 3 Time Lag	na	na	na	na	2.7712	2.20
Covariance Alternative 4 Time Lag	na	na	na	na	2.2793	5.00
Mean Value of Time ($/hour)	11.22		15.49		18.7291	
Std. Deviation Value of Time ($/hour)	na		24.59		14.8155	

Note: na = not applicable.

captured. This is similar to the comparison between VOT for route-type choice only and the nested model, which captured both mode and route-type choice. This suggests that much of the variance associated with VOT across a population may be due to not capturing other choice dimensions, in addition to inherent taste variation across users.

Other Choice Dimensions

The choice framework described above, which includes such dimensions as TOD, mode, car occupancy, and route type, can also be effectively employed to incorporate congestion and pricing effects on all other choice dimensions, including destination choice, tour and trip frequency, daily activity patterns, and car ownership. This technique, which is based on various derived accessibility measures, has been already successfully employed in many ABMs in practice. The advantage of using accessibility measures is that all LOS variables, including travel time reliability measures included in the route, mode, or TOD utility components, will be automatically incorporated in all upper-level choice models that include these accessibility measures as explanatory variables. This technique, however, does not preclude using some relevant congestion, pricing, and reliability effects in the upper-level choice model directly. This

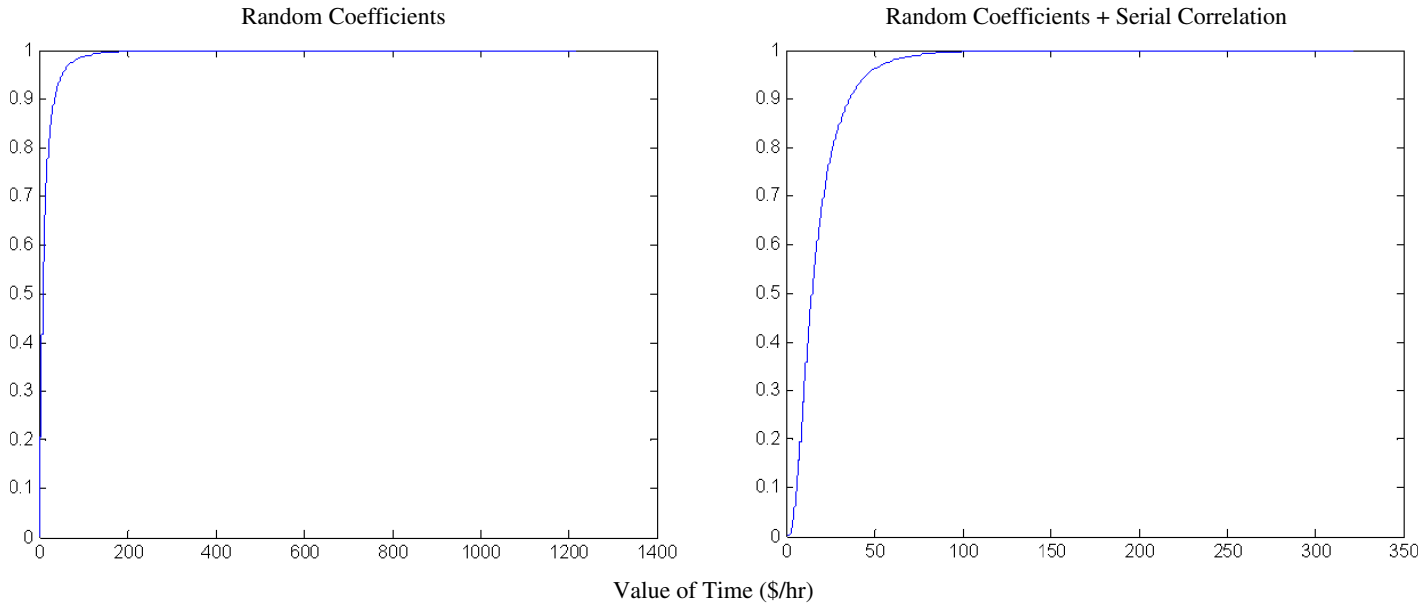

Figure 3.11. Cumulative distribution functions for VOT for* (left) *random coefficients and* (right) *random coefficients and serial correlation.

direction has been less explored and represents a sound possible topic for future research. In the current report, the team further describes the approach based on accessibility measures derived from the lower-level tour and trip models that are immediately implementable with the highway utility (generalized cost) functions described above.

General Forms of Accessibility Measures

Multiple accessibility measures have been applied in the recently developed ABMs for such metropolitan regions as Sacramento and San Diego, California; and Phoenix, Arizona. Most of the applied accessibility measures represent simplified destination choice logsums, which is the composite utility of travel across all modes to all potential destinations from an origin zone to all destination zones in different TOD periods. This way the accessibility measure is essentially a zonal characteristic that can be stored as a vector indexed by a traffic analysis zone (TAZ). Another type of accessibility measure that is calculated in the process of calculations for the zonal measure is the measure of impedance between the zones. Accessibilities of this type have to be stored as TAZ-to-TAZ matrices.

These accessibility measures are primarily needed to ensure that the upper-level models in the ABM hierarchy, such as car ownership, daily activity pattern, and (nonmandatory) tour frequency, are sensitive to improvements of transportation LOS across all modes, as well as changes in land use. Accessibility measures are similar in nature to density measures and can be thought of as continuously buffered "fuzzy" densities.

Accessibility measures are needed because it is infeasible to link all choices by full logsums due to the number of potential alternatives across all dimensions (activities, modes, time periods, tour patterns, and daily activity patterns). Accessibility measures reflect the opportunities to implement a travel tour for a certain purpose from a certain origin (residential or workplace). They are used as explanatory variables in the upper-level models (daily activity pattern type and tour frequency), and the corresponding coefficients are estimated along with the coefficients for person and household variables.

The Sacramento, Phoenix, and San Diego ABMs are among the first advanced travel models that completely avoid the "flat" area-type dummies (such as CBD, urban, suburban, and rural dummies) that are frequently used in other models to explain such choice as car ownership, tour or trip frequency, and mode choice. These qualitative labels have been completely replaced by the physical measures of accessibility sensitive to travel time, cost, (and potentially) reliability.

The applied zonal accessibility measures have the following general form:

$$A_i = \ln\left[\sum_{j=1}^{I} S_j \times \exp\left(TMLS_{ij}\right)\right] \tag{3.40}$$

where

$i, j \in I$ = origin and destination zones;

A_i = accessibility measure calculated for each origin zone;

S_j = attraction size variable for each potential destination zone; and

$TMLS_{ij}$ = TOD and mode choice logsum as the measure of impedance.

The composite travel impedance between zones, which can be referred to as an *O-D accessibility measure,* is calculated as a two-level logsum taken over the TOD periods and modes:

$$TMLS_{ij} = \mu \ln\left[\sum_{t=1}^{2} \exp\left(MLS_{tij} + \alpha_t\right)\right] \tag{3.41}$$

where

$t = 1, 2 =$ TOD periods (currently peak and off-peak are used);

$MLS_{tij} =$ mode choice logsum for a particular TOD period;

$\alpha_t =$ TOD-specific constant; and

$\mu =$ nesting coefficient for mode choice under TOD choice.

In this form, the destination choice accessibility measure is essentially a sum of all attractions in the region discounted by the travel impedance. Note that this measure is sensitive to travel improvements in both peak and off-peak periods. The relative impact of each period is regulated by the TOD-specific constant that is estimated for each travel segment (or activity type).

Accessibility measures are linearly included in a utility function of an upper-level model. To preserve consistency with the random-utility choice theory, the coefficient for any accessibility measure should be between zero and one, although it is not as restrictive as in a case of a proper nested logit model.

The general logic of inclusion of accessibility measures in travel models is as follows. For models that generate activity patterns, tours, and trips for which specific destinations are not known yet, zonal accessibility measures should be applied that describe the density of the supply of potential activity locations. For models in which accessibility to an already known location (modeled prior in the model chain) is evaluated, O-D measures should be used. In this case, there is no need for a size variable.

Size Variables by Activity Type

Size variables are prepared for each TAZ and segmented by activity type (trip purpose). The zonal size variables are calculated as linear combinations of the relevant land use variables. The corresponding coefficients can be preestimated by means of regressions of the expanded observed trip ends on the available land use variable, primarily employment types. In this sense, the size variables are similar to conventional trip attraction models.

A more theoretically consistent but also more complicated procedure would involve a simultaneous estimation of the size terms and impedance functions in the destination choice context by Equation 3.40. The estimation results for all activity types with the Phoenix data are presented in Table 3.13 for nonwork purposes (numbered from 4 through 9), a special reserved at-work subtours purpose (10), and all home-based nonwork (nonmandatory) purposes (by a combined nonwork attraction measure, 11). The explanatory variables in the rows are referred to by the tokens used in the model application, in which "nxx" implies employment for the North American Industry Classification System (NAICS) code "xx." The resulting size variables in the columns are designated by the purpose number and an abbreviation of the purpose.

Table 3.13. Zonal Size Variables for Accessibility Measures by Activity Type

Explanatory Variable		Size Variable by Activity Type							
Variable	**Description**	**4 = escort**	**5 = shop**	**6 = maint**	**7 = eating**	**8 = visit**	**9 = discr**	**10 = at work**	**11 = all**
Total_HH	Total number of households	1.0000	na	na	na	0.1421	0.3595	na	0.5016
retail	Retail employment (*n*44 + *n*45)	na	4.2810	1.4185	1.2908	na	0.4387	0.5403	7.4291
n51	Information	na	na	0.7091	na	na	na	na	0.7091
n52	Finance and insurance	na	na	na	na	na	na	0.1265	na
n53	Real estate rental leasing	na	na	2.4753	na	na	na	na	2.4753
n55	Management of companies and enterprises	na	na	na	na	na	na	1.3759	na
n56	Administrative and support	na	na	na	na	na	na	0.2357	na
n62	Health care, social assistance	na	na	1.0618	na	0.2349	na	na	1.2968
n71	Arts, entertainment, recreation	na	na	na	0.3224	na	0.9049	na	1.2273
n72	Accommodation, food services	na	1.1224	na	1.0458	na	0.4422	0.2809	2.6104
n92	Public administration	na	na	0.5356	na	na	na	0.2265	0.5356
total_emp	Total employment	na	na	na	na	na	na	0.1578	na

Note: All = combined nonwork attraction measure; at work = at-work subtours purpose; discr = discretionary purpose; eating = eating-out purpose; escort = escorting purpose; maint = (household) maintenance purpose; shop = shopping purpose; visit = visiting relatives and friends purpose.

For escorting purpose (purpose = 4), the size variable is set to the total population. This is a special purpose for which accessibility to a potential destination does not directly relate to the household decision to escort one of the household members (most frequently a child). Also, despite the fact that escorting is most frequently associated with the school purpose for the escorted person (child), the density of schools around the respondent's residence does not mean that escorting would occur more frequently. On the contrary, if a child can walk to the nearby school escorting will not be needed. Population density (accessibility to population can be viewed as continuously buffered population density) is somewhat the most reasonable zonal size variable that affects probability of escorting (all else being equal, meaning the household composition and necessity of escorting). Population density is the only accessibility measure for which both negative and positive signs can be accepted in the tour and activity frequency model. All other accessibility measures are accepted only if they have a logical positive sign.

For shopping purpose (purpose = 5), the main attractions are logically associated with retail employment and food services. Food services are frequently intertwined with shopping and it is difficult to completely separate these two land use types. It is equally true for both major shopping malls and small street shops or restaurants. It is recommended that in the future shopping size variables should be enriched with such explanatory variables as floor area to better distinguish between large shopping malls and small street shops. The (household) maintenance purpose (purpose = 6) includes a range of activities, such as personal business, banking, and visiting a post office, doctor, dentist, or lawyer, and is scattered over a wide range of related employment types including retail, information, real estate, rental, leasing, health care, social assistance, and public administration.

Eating out (purpose = 7) and discretionary (purpose = 9) purposes are closely intertwined and frequently combined in the same tour. They share the same attraction variables that relate to retail employment, recreation and entertainment, and food services, although the coefficients are logically different. In addition, discretionary purpose includes population as an additional attraction factor that serves as a proxy for such factors as sport facilities and playing grounds. It is recommended that in the future nonemployment variables like land or public parks and floor areas for sport facilities should be added, as these would enrich the attraction model for discretionary activities.

The purpose visiting relatives and friends (purpose = 8) is a special purpose for which the major attraction factor is population (number of households). In addition, visiting frequently occurs at a hospital, which is measured by employment in health. Attraction factors for trips that originated from the workplace (purpose = 10) include many variables, because at-work travel comprises three main purposes. First, at-work travel includes eating out during the lunch break, which is reflected in such attractions as retail employment and food services. Second, it may include business trips for meetings, which is reflected in such employment categories as management of companies and administration (the most probable places for business meetings) and some proportion of total employment. Third, workers might use the lunch break for personal business and shopping, which is reflected in such employment categories as finance, insurance, and public administration. Finally, a size variable that expresses total attractions for all nonmandatory home-based purposes (purposes 4–9) includes a mix of all corresponding employment types and population. Logically, retail employment plays a major role in this mix.

In addition to the complex size variables for nonmandatory activities, an ABM requires several size variables for zonal accessibility measures to mandatory activities. These are primarily used in the choice models for work from home and schooling from home. These size variables are simpler because they include all relevant variables with a coefficient of 1.0. For work from home, the size variable is employment for the relevant occupation (divided into the five categories used in the 2008 National Household Travel Survey used to estimate the Phoenix ABM). These five categories are related to employment by the NAICS codes used as the source of explanatory variables. For schooling from home, the size variable is enrollment in the corresponding school type broken into three categories: grades K–8 (elementary or middle school), grades 9–12 (high school), and university or college. The corresponding size variables are summarized in Table 3.14.

Impedance Functions by Person, Household, and Activity Type

Impedance functions are calculated as O-D matrices of logsums over modes and TOD periods (peak and off peak) according to Equation 3.41. The calculation is based on mode choice utilities that have to be calculated for all modes and TOD periods as the first step. These utilities are then combined in the composite logsum at the second step. Both steps are described below.

Mode Utilities

For calculation of accessibility measures, the set of modes is simplified to include five main modes: 1 = SOV, 2 = HOV, 3 = walk to transit (WT), 4 = drive to transit (DT), and 5 = nonmotorized (NM). The WT and DT utilities are based on the best transit skims implemented for the entire transit network including all modes. Mode utilities are also calculated separately for each of the four aggregate travel purposes: 1 = work, 2 = university, 3 = school, and 4 = other. Segmentation by travel purpose is essential because each travel purpose is characterized by a different set of mode preferences. For example, DT

Table 3.14. Zonal Size Variables for Mandatory Activities

Explanatory Variable		Size Variable	
Variable	**Description**	**Variable**	**Description**
n42	Wholesale trade	p12_whom1	Sales or marketing
n52	Finance and insurance		
n44	Retail trade	p13_whom2	Clerical, administrative, or retail
n45	Retail trade		
n53	Real estate and rental and leasing		
n71	Arts, entertainment, and recreation		
n72	Accommodation and food services		
n92	Public administration		
n11	Agriculture, forestry, fishing, hunting	p14_whom3	Production, construction, manufacturing, or transport
n21	Mining, quarrying, oil and gas extraction		
n22	Utilities		
n23	Construction		
n31	Manufacturing		
n32	Manufacturing		
n33	Manufacturing		
n48	Transportation and warehousing		
n49	Transportation and warehousing		
n51	Information	p15_whom4	Professional, managerial, or technical
n54	Professional, scientific, and technical services		
n55	Management of companies and enterprises		
n56	Administrative and support and waste management and remediation services		
n61	Educational services		
n62	Health care and social assistance		
n81	Other services (except public administration)	p16_whom5	Person care and services
Enroll1	Enrollment K–8	p17_shom1	Enrollment primary and middle
Enroll2	Enrollment 9–12	p18_shom2	Enrollment high school
Enroll3	Enrollment university and college	p19_shom3	Enrollment university and college

is frequently chosen for work purpose, but it is practically not observed for the purposes of school trips or other trips. All nonwork purposes are aggregated for calculation of impedances, although they are separated with respect to size variables. Additional important segmentation relates to household car sufficiency. The team distinguishes at this stage between three household groups: 1 = household without cars, 2 = household with fewer cars than workers, and 3 = households with number of cars greater than or equal to number of workers. This distinction is important because car sufficiency strongly affects mode availability and preferences.

Overall, by combining five aggregate modes with four travel purposes, three car-sufficiency groups, and two TOD periods, a set of $5 \times 4 \times 3 \times 2 = 120$ mode utilities was precalculated for all O-D pairs. The components of the mode utility functions and corresponding coefficients are summarized in Table 3.15. The coefficients shown were adopted for the San Diego and Phoenix ABMs. All coefficients are generic across TOD periods. The distinction between peak and off-peak utilities is due to different LOS variables. Mode utilities can incorporate STD, perceived highway time, or any other measure of travel time reliability if is supported by the network simulation and skimming procedures. This is not currently the case with the ABMs in practice, primarily because of the lack of effective network procedures that could generate reliability measures. This issue is the focus of such SHRP 2 projects as L04 and C10, which are currently under way. However, the current research lays down a complete methodology for incorporating reliability in travel demand models that is fully compatible with the potential network procedures.

Table 3.15. Components and Coefficients of Mode Utilities

Variable	SOV	HOV	WT	DT	NM
Work Travel Purpose					
SOV time (min)	–0.03	na	na	na	na
HOV time (min)		–0.03	na	na	na
Highway distance (mi)	–0.015	–0.01	na	na	–1.5
Highway distance >3 mi, dummy	na	na	na	na	–999
WT weighted time (min)	na	na	–0.03	na	na
WT fare (cents)	na	na	–0.002	na	na
WT in-vehicle time <1 min, dummy	na	na	–999	na	na
DT weighted time (min)	na	na	na	–0.03	na
DT fare (cents)	na	na	na	–0.002	na
DT in-vehicle time <1 min, dummy	na	na	na	–999	na
Zero-car household	–999	–3.0	na	na	na
Cars fewer than workers	–1.5	–2.0	na	na	na
Cars greater than or equal to workers		–2.5	na	na	na
University Travel Purpose					
SOV time (min)	–0.03	na	na	na	na
HOV time (min)	na	–0.03	na	na	na
Highway distance (mi)	–0.03	–0.02	na	na	–1.5
Highway distance >3 mi, dummy	na	na	na	na	–999
WT weighted time (min)[a]	na	na	–0.03	na	na
WT fare (cents)	na	na	–0.004	na	na
WT in-vehicle time <1 min, dummy	na	na	–999	na	na
DT weighted time (min)[b]	na	na	na	–0.03	na
DT fare (cents)	na	na	na	–0.004	na
DT in-vehicle time <1 min, dummy	na	na	na	–999	na
Zero-car household	–999	–2.0	na	na	na
Cars fewer than workers	–1.5	–1.0	na	na	na
Cars greater than or equal to workers	0	–1.5	na	na	na
School Travel Purpose					
SOV time (min)	–0.05	na	na	na	na
HOV time (min)	na	–0.05	na	na	na
Highway distance (mi)	–0.06	–0.04	na	na	–1.5
Highway distance >3 mi, dummy	na	na	na	na	–999
WT weighted time (min)[a]	na	na	–0.03	na	na
WT fare (cents)	na	na	–0.006	na	na
WT in-vehicle time <1 min, dummy	na	na	–999	na	na
DT weighted time (min)[b]	na	na	na	–0.03	na
DT fare (cents)	na	na	na	–0.004	na
DT in-vehicle time <1 min, dummy	na	na	na	–999	na
Zero-car household	–999	–1.0	na	–5.0	2.0
Cars fewer than workers	–1.5	0	na	–5.0	2.0
Cars greater than or equal to workers	0	–0.5	na	–5.0	2.0

(continued on next page)

Table 3.15. Components and Coefficients of Mode Utilities (continued)

Variable	SOV	HOV	WT	DT	NM
Other Travel Purpose					
SOV time (min)	–0.03	na	na	na	na
HOV time (min)	na	–0.03	na	na	na
Highway distance (mi)	–0.03	–0.02	na	na	–1.5
Highway distance >3 mi, dummy	na	na	na	na	–999
WT weighted time (min)[a]	na	na	–0.03	na	na
WT fare (cents)	na	na	–0.004	na	na
WT in-vehicle time <1 min, dummy	na	na	–999	na	na
DT weighted time (min)[b]	na	na	na	–0.03	na
DT fare (cents)	na	na	na	–0.004	na
DT in-vehicle time <1 min, dummy	na	na	na	–999	na
Zero-car household	–999	–3.0	na	–5.0	na
Cars fewer than workers	–1.5	–2.0	na	–5.0	na
Cars greater than or equal to workers	0	–2.5	na	–5.0	na

Note: SOV = Single Occupancy Vehicle; HOV = High Occupancy Vehicle; WT = Transit with Walk access; DT = Transit with Drive Access; NM = Non-motorized modes; and na = not applicable.

[a] WT weighted time includes in-vehicle time and out-of-vehicle time with weight = 2.5. Out-of-vehicle time includes initial wait, transfer wait, access walk, transfer walk, egress walk, and a 4-minute penalty for each transfer.

[b] DT weighted time additionally includes access drive in out-of-vehicle time.

Mode and TOD Choice Logsums

After mode utilities have been calculated for each mode, purpose, car-sufficiency group, and TOD period, they are combined into composite O-D accessibility measures; that is, mode and TOD choice logsums are derived by Equation 3.41. The list of logsum measures that have to be prepared to support various accessibility measures is summarized in Table 3.16.

Overall, 21 O-D accessibility measures are prepared to support the various zonal accessibility measures needed for different submodels of the MAG ABM. The structure of each logsum and associated parameters are summarized in Table 3.17. This table essentially represents a control file for the impedance (O-D) part of the program that calculates accessibility measures.

Each impedance measure is associated with a certain aggregate travel purpose (1–4) for which the mode utilities are calculated according to the coefficients in Table 3.15. Depending on the type of accessibility measure, car sufficiency is then taken into account. If a general accessibility measure is calculated that will be applied in the model system before the car-ownership model, the mode utilities are averaged across all car-sufficiency groups with the weight that reflects the observed proportion between different car-sufficiency groups in the region. If an accessibility measure is calculated for a specific car-sufficiency group (i.e., it will be applied after the car-ownership model), then the mode utilities for this specific group are used.

Not every mode is included in each logsum. The set of modes is restricted for two reasons. First, some modes are not observed for some of the trip purposes. For example, DT is relevant for work trips only. Second, certain modes are made unavailable in order to calculate a specific (mode-restricted) type of accessibility needed for a particular behavioral model. For example, mode-specific accessibilities that are used in the car-ownership model are based on a single representative mode each. Accessibilities that describe individual activities should logically exclude HOV. Accessibilities that describe joint activities naturally exclude SOV. Accessibilities that describe auto dependency include only modes that need an auto (SOV, HOV, and DT). Accessibilities that describe auto nondependency include only modes that do not need an auto (WT and NM).

Finally, to complete the logsum calculation across TOD periods, a bias constant for off-peak periods is specified (the peak period is used as the reference alternative with zero bias). This constant is set to replicate the observed proportion of trips in the peak period versus the off-peak period.

List of Zonal Accessibility Measures Adopted for Advanced Activity-Based Models

The set of zonal accessibility measures incorporated in the Sacramento, San Diego, and Phoenix ABMs (with some simplifications to create a common denominator across different models) is summarized in Table 3.18. The variety of measures stems from the combination of different size variables segmented by

Table 3.16. List of Mode and TOD Choice Logsums

Impedance	Accessibility from the Given (Residential) Zone to	Token
1	Workplace by all modes for all car-sufficiency groups	Work
2	University by all modes for all car-sufficiency groups	Univ
3	School by all modes for all car-sufficiency groups	Scho
4	Nonmandatory activity location by auto	Auto
5	Nonmandatory activity location by WT	Tran
6	Nonmandatory activity location by NM (walk)	Nonm
7	Nonmandatory activity by all modes, individual travel, zero-car household	Indi_0
8	Nonmandatory activity by all modes, individual travel, cars < workers	Indi_1
9	Nonmandatory activity by all modes, individual travel, cars ≥ workers	Indi_2
10	Nonmandatory activity by all modes, joint travel, zero-car household	Join_0
11	Nonmandatory activity by all modes, joint travel, cars < workers	Join_1
12	Nonmandatory activity by all modes, joint travel, cars ≥ workers	Join_2
13	Escort accessibility, joint travel, zero-car household	Esco_0
14	Escort accessibility, joint travel, cars < workers	Esco_1
15	Escort accessibility, joint travel, cars ≥ workers	Esco_2
16	Workplace by auto modes for all car-sufficiency groups (auto dependency)	Wrkad
17	University by auto modes for all car-sufficiency groups (auto dependency)	Unvad
18	School by auto modes for all car-sufficiency groups (auto dependency)	Schad
19	Workplace by nonauto modes (nonauto dependency)	Wrknad
20	University by nonauto modes (nonauto dependency)	Unvnad
21	School by nonauto modes (nonauto dependency)	Schnad

Table 3.17. Structure of Mode and TOD Choice Logsums

Token	Purpose	Car Sufficiency			Modes Included					Off-Peak Constant
		Zero Cars	Cars < Workers	Cars ≥ to Workers	SOV	HOV	WT	DT	NM	
Work	1 = Work	0.05	0.35	0.6	1	1	1	1	1	–0.9
Univ	2 = Univ	0.05	0.35	0.6	1	1	1	na	1	–0.5
Scho	3 = Scho	0.05	0.35	0.6	1	1	1	na	1	–1.2
Auto	4 = Other	0.05	0.35	0.6	1	na	na	na	na	0.5
Tran	4 = Other	0.05	0.35	0.6	na	na	1	na	na	0.5
Nonm	4 = Other	0.05	0.35	0.6	na	na	na	na	1	0.5
Indi_0	4 = Other	1	na	na	1	na	1	na	1	0.5
Indi_1	4 = Other	na	1	na	1	na	1	na	1	0.5
Indi_2	4 = Other	na	na	1	1	na	1	na	1	0.5
Join_0	4 = Other	1	na	na	na	1	1	na	1	0.5
Join_1	4 = Other	na	1	na	na	1	1	na	1	0.5
Join_2	4 = Other	na	na	1	na	1	1	na	1	0.5
Esco_0	4 = Other	1	na	na	na	1	na	na	1	–0.5
Esco_1	4 = Other	na	1	na	na	1	na	na	1	–0.5
Esco_2	4 = Other	na	na	1	na	1	na	na	1	–0.5
Wrkad	1 = Work	0.05	0.35	0.6	1	1	na	1	na	–0.9
Unvad	2 = Univ	0.05	0.35	0.6	1	1	na	1	na	–0.5
Schad	3 = Scho	0.05	0.35	0.6	1	1	na	1	na	–1.2
Wrknad	1 = Work	0.05	0.35	0.6	na	na	1	na	1	–0.9
Unvnad	2 = Univ	0.05	0.35	0.6	na	na	1	na	1	–0.5
Schnad	3 = Scho	0.05	0.35	0.6	na	na	1	na	1	–1.2

Table 3.18. Zonal Accessibility Measures

Measure	Size Variable		Impedance Measure		Model in Which Applied
	No.	Token	No.	Token	
1	12	Whom1	1	Work	Work from home
2	13	Whom2	1	Work	Work from home
3	14	Whom3	1	Work	Work from home
4	15	Whom4	1	Work	Work from home
5	16	Whom5	1	Work	Work from home
6	17	Shom1	3	Scho	Schooling from home
7	18	Shom2	3	Scho	Schooling from home
8	19	Shom3	2	Univ	Schooling from home
9	11	AllNM	4	Auto	Car ownership
10	11	AllNM	5	Tran	Car ownership
11	11	AllNM	6	Nonm	Car ownership
12	11	AllNM	7	Indi_0	Coordinated daily activity–travel pattern
13	11	AllNM	8	Indi_1	Coordinated daily activity–travel pattern
14	11	AllNM	9	Indi_2	Coordinated daily activity–travel pattern
15	11	AllNM	10	Join_0	Coordinated daily activity–travel pattern
16	11	AllNM	11	Join_1	Coordinated daily activity–travel pattern
17	11	AllNM	12	Join_2	Coordinated daily activity–travel pattern
18	5	Shop	10	Join_0	Joint tour frequency
19	5	Shop	11	Join_1	Joint tour frequency
20	5	Shop	12	Join_2	Joint tour frequency
21	6	Main	10	Join_0	Joint tour frequency
22	6	Main	11	Join_1	Joint tour frequency
23	6	Main	12	Join_2	Joint tour frequency
24	7	Eati	10	Join_0	Joint tour frequency
25	7	Eati	11	Join_1	Joint tour frequency
26	7	Eati	12	Join_2	Joint tour frequency
27	8	Visi	10	Join_0	Joint tour frequency
28	8	Visi	11	Join_1	Joint tour frequency
29	8	Visi	12	Join_2	Joint tour frequency
30	9	Disc	10	Join_0	Joint tour frequency
31	9	Disc	11	Join_1	Joint tour frequency
32	9	Disc	12	Join_2	Joint tour frequency
33	4	Esco	13	Esco_0	Allocated tour frequency
34	4	Esco	14	Esco_1	Allocated tour frequency
35	4	Esco	15	Esco_2	Allocated tour frequency
36	5	Shop	7	Indi_0	Allocated tour frequency
37	5	Shop	8	Indi_1	Allocated tour frequency
38	5	Shop	9	Indi_2	Allocated tour frequency
39	6	Main	7	Indi_0	Allocated tour frequency
40	6	Main	8	Indi_1	Allocated tour frequency
41	6	Main	9	Indi_2	Allocated tour frequency
42	7	Eati	7	Indi_0	Individual tour frequency

(continued on next page)

Table 3.18. Zonal Accessibility Measures (continued)

Measure	Size Variable		Impedance Measure		Model in Which Applied
	No.	Token	No.	Token	
43	7	Eati	8	Indi_1	Individual tour frequency
44	7	Eati	9	Indi_2	Individual tour frequency
45	8	Visi	7	Indi_0	Individual tour frequency
46	8	Visi	8	Indi_1	Individual tour frequency
47	8	Visi	9	Indi_2	Individual tour frequency
48	9	Disc	7	Indi_0	Individual tour frequency
49	9	Disc	8	Indi_1	Individual tour frequency
50	9	Disc	9	Indi_2	Individual tour frequency
51	10	Atwo	7	Indi_0	Individual subtour frequency
52	10	Atwo	9	Indi_2	Individual subtour frequency

the underlying activity type with different impedance measures segmented by trip purpose and person and household type. Models such as car ownership (mobility attributes), work and schooling from home, and coordinated daily activity–travel pattern are very good illustrations for zonal accessibility measures with some components that relate to O-D accessibility measures. Models such as usual workplace and school location are based on O-D accessibility measures.

The 52 zonal accessibility measures combine 19 size variables (numbered and tokenized in Tables 3.13 and 3.14) and 15 impedance measures (numbered and tokenized in Tables 3.16 and 3.17). There are six impedance measures (16–21) that are used only as O-D accessibilities. Multiple examples of impacts of the accessibility measures on different aspects of travel behavior can be found in model estimation reports for the Sacramento, San Diego, and Phoenix ABMs.

CHAPTER 4

Network Simulation Procedures to Support Congestion and Pricing Studies

This chapter discusses approaches for integrating models of user behavior in network modeling and simulation frameworks to support analysis and evaluation of pricing and other congestion-related measures. It also showcases the application of such integrated demand–network simulation procedures in an actual large-scale regional network: the New York City (NYC) best practice regional network. To the team's knowledge, the network considered in this application is the largest network (in terms of number of links and nodes, as well as simulated vehicles and trip makers) to which a simulation-based dynamic traffic assignment (DTA) procedure has been applied. The three major sections of this chapter present (1) a general overview and recommended methods of network simulation modeling, with a focus on an integrated multidimensional network choice model framework; (2) a demonstration of the proposed model framework with an application to the NYC regional network to support congestion and pricing studies; and (3) a summary of network simulation. Mathematical formulations and solution algorithms for the proposed integrated model are presented in greater detail in Appendix A. Details of the calibration procedure and its application for the estimation of time-dependent origin–destination (O-D) demand with multiple vehicle types are provided in Appendix A.

General Review and Recommended Methods of Network Simulation

The review of and recommendations for network simulation are divided into four subsections as follows:

- Summary of the challenges of integrating user decisions in network simulations models;
- Presentation of an integrated multidimensional network choice model framework to support congestion and pricing studies;
- Presentation of a simulation-based column generation solution framework to solve the proposed problem; and
- Presentation of algorithmic procedures and discussion of associated challenges for applying the proposed solution method to large-scale regional networks.

Integrating User Decisions in Network Simulation Models: Summary of Challenges

A regional transportation model is a mathematical representation of travel demand and network supply in a metropolitan area. The travel demand side is a result of interactions among various economic and social activities in the region. The network supply side generally includes a highway network and a transit network for the region. The highway network consists of arterials, freeways, and toll roads; the transit network represents all public transportation modes, such as buses, ferries, and trains. Generally two types of travel demand models, trip-based models and activity-based models (ABMs), represent travel activities and choices of travel destinations, frequency of travel, mode, and so forth. The network supply model represents how a given travel demand is distributed and propagated through a transportation network, namely traffic assignment, which essentially assigns the travel demand (i.e., trips or activity chains) to the transportation network links and determines the corresponding service levels of the transportation network elements.

Network models used in practice have typically followed static assignment procedures, which assume that traffic flows and associated trip times are constant over time. Because these assumptions do not capture observed temporal patterns of congestion build up and dissipation in actual networks, analysts have moved toward DTA procedures. Static traffic assignment models typically rely on analytical link volume-delay functions (e.g., the Bureau of Public Roads function) to capture the dependence of level of service (LOS) on flow levels. In DTA models, traffic simulation procedures have

increasingly been used to realistically capture traffic dynamics in practice.

Most state departments of transportation (DOTs) and metropolitan planning organizations (MPOs) rely on static regional transportation models that have often been developed over several decades. In general, these static models execute travel demand and network supply models separately, in a sequential process, or introduce some elementary feedback loop between them. It has long been recognized that the travel demand side (especially frequency of travel, mode choice, and so forth) is influenced by the network supply side. From a realistic behavioral standpoint, integrating travel demand (especially for mode choice) with network supply (traffic assignment model) is required to address a wide range of transportation options and demand management policies (e.g., congestion pricing) in a regional large-scale transportation system.

As noted, conventional static assignment models assume stationary traffic states, in conjunction with simplified route choice assumptions and link volume-delay functions. As such, their applicability is severely limited for the evaluation of dynamic transportation management policies (e.g., dynamic pricing) over relatively short time scales (e.g., 5 or 15 minutes). DTA models with an underlying network traffic simulator provide a more realistic representation of traffic in terms of congestion, queues, and dynamic route choice than static traffic assignment models, especially for travelers' choices of using toll versus nontoll roads over time.

Simulation-based DTA models have been successfully applied to conduct analyses of pricing, reliability, and congestion at a corridor level. For example, such analyses were part of an integrated corridor management tool applied to the CHART network between Baltimore, Maryland, and Washington, D.C. (Zhou et al. 2008; National Cooperative Highway Research Program 2009) using a simplified multinomial logit modal choice model. However, corridor models are not intended as a substitute for regional models, but as an effective complementary tool to achieve a more detailed level of analysis. The corridor models are useful for a detailed analysis of a route under congestion, and specifically for a better representation of facility-level choices between managed lanes and general-purpose lanes, as well as issues such as queuing phenomena and congestion at toll plazas. Decision-support tools to evaluate congestion-pricing policies call for a regional-level integrated user decision and DTA model incorporating multidimensional network choice behavior. In particular, since most current travel demand models used by state DOTs or MPOs (e.g., modal choice models) are rather complex due to the large network size and the presence of multiple modal alternatives at the regional level, this combined model requires a seamlessly and correctly integrated modeling framework to connect DTA models with available well-calibrated travel demand models.

Accordingly, this study aims to demonstrate how to integrate a travel demand model with a simulation-based DTA model to support congestion and pricing studies in a large-scale regional transportation network using the NYC regional transportation network as an extended example. The study provides a foundational framework that represents an evolution from current practice toward a conceptually, theoretically, and methodologically sound approach to address heterogeneous user responses to congestion, pricing, and reliability in large-scale regional multimodal transportation networks.

The integration of travel demand and dynamic network simulation models to support congestion and pricing studies in large-scale regional transportation networks gives rise to several challenges, including the following:

- Capturing user responses to congestion, pricing, and reliability is best accomplished through microsimulation of individual traveler decisions in a network platform. These responses must be considered in a network setting, not at the facility level, and the time dimension is essential to evaluating the impact of congestion pricing and related measures. Hence, a dynamic analysis tool is required;
- Incorporating heterogeneity of user preferences is an essential requirement for modeling user responses to pricing in a network setting, as discussed in Chapter 3. New algorithms that exploit parametric multicriteria shortest-path procedures allow travelers' value of time (VOT), which determines users' choice of path and mode in response to prices, to be continuously distributed across the population of travelers. Efficient implementations of these algorithms have been demonstrated for large network applications for the first time as part of this study;
- Simulation-based DTA models have gained considerable acceptance in the past few years, yet adoption in practice remains in its infancy, especially for large-scale regional transportation networks. The current generation of available models only considers fixed, albeit time-varying, O-D trip patterns. Greater use and utility will result from consideration of a more complete set of travel choice dimensions and incorporation of user attributes, including systematic and random heterogeneity of user preferences;
- Algorithms for finding equilibrium time-varying flows have been based on the relatively inefficient method of successive averages. Its implementation in a flow-based procedure did not scale particularly well for application to large metropolitan networks. New implementations of the method of successive averages and other algorithms that exploit the particle-based approach of DTA simulators have been proposed and demonstrated on large-scale regional transportation networks (Sbayti et al. 2007; Lu et al. 2009); and
- Regional networks are large-scale applications of network models and require substantial computational time and

memory storage for the various algorithmic components of these procedures, especially shortest-path calculation, traffic simulation, and traffic assignment. In addition, large-scale regional networks require large amounts of memory to store data for path calculation and traffic simulation, as well as traffic assignment.

The next section presents an integrated model framework to evaluate pricing and reliability to overcome the abovementioned challenges. The proposed integrated model framework is a demonstration of a trip-based integration of a well-calibrated modal choice model in practice and a simulation-based dynamic traffic microassignment model. However, this framework is sufficiently flexible to incorporate other dimensions than modal choice, such as destination choice and departure time choice, from the demand side. In addition, this framework can be readily extended to an activity-based integration of demand models and an activity-based dynamic traffic microassignment model.

Integrated Model Framework to Evaluate Pricing and Reliability

Problem Statement and Assumptions

The starting point is a network of links and nodes representing the study area and a population of travelers with desired origins, destinations, and activity times reflecting their daily activity schedules. The problem considered here is downstream of the activity-scheduling process, whereby activity patterns have been mapped onto trips with known origins, destinations, and departure times. However, it is also possible to define the problem at the daily activity pattern formation or choice level with full integration of the supply-modeling procedures with the higher-level activity choices. The basic methodological framework presented here could be readily expanded to accommodate such integration, albeit with richer path-finding procedures in the combined travel activity time–space network and heavier computational burden. The focus here is limited to demonstrating practical procedures to achieve integration of a rich multimodal path- and mode-choice process in the network-modeling process.

A more formal statement of the problem and key assumptions is as follows. Given a time-varying network $G = (N, A)$, where N is a finite set of nodes and A is a finite set of directed links, the time period of interest (planning horizon) is discretized into a set of small time intervals, $H = \{t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots, t_0 + T\Delta t\}$. Here t_0 is the earliest possible departure time from any origin node, Δt is a small time interval during which no perceptible changes in either traffic conditions or travel cost (or both) occur, and T is a large number such that the intervals from t_0 to $t_0 + T\Delta t$ cover the planning horizon H. The time-dependent zonal demand q^{wt} over the study horizon represents the number of individual travelers of an O-D pair w ($w \in W$) at departure time t ($t \in T$). The set of available modes is denoted as M. The integrated model in this study is designed to find a dynamic network equilibrium mode–path flow pattern by recognizing multiple dimensions of network choice behavior (i.e., mode choice decision, highway user heterogeneity, and reliability of route choice). Essentially, this is an integrated dynamic traveler mode–path assignment problem on a multimodal transportation network.

Recognizing Dynamic Mode Choice Decision

Associated with each mode m is the mode flow y_m^{wt}, $\forall m \in M$, $w \in W$, $t \in T$ and corresponding mode choice probability $p_m^{wt}(y)$, $\forall m \in M$, $w \in W$, $t \in T$, where $y = \{y_m^{wt} | \forall m \in M, w \in W, t \in T\}$ is the mode flow vector for all O-D pairs and departure times. A key behavioral assumption for the mode choice decision is as follows: in a random utility maximization framework, each traveler chooses a mode that maximizes his or her perceived utility. With no loss of generality, the choice probability of each mode $p_m^{wt}(y)$, $\forall m \in M$, $w \in W$, $t \in T$ can be determined as follows:

$$p_m^{wt}(y) = \Pr\left[U_m^{wt}(y) = \max_{m \in M}\{U_m^{wt}(y)\}\right], \quad \forall m \in M, w \in W, t \in T \qquad (4.1)$$

where $U_m^{wt}(y)$ is a utility function of mode m and $\Pr[\bullet]$ is a choice probability function. Note that the utility function U_m^{wt} is a function of travelers' characteristics, mode attributes, and a random term that determines the structure of the choice model. The exact form of $\Pr[\bullet]$ is defined by the underlying random error structure of the choice model.

Daganzo and Sheffi (1977) defined the static stochastic user equilibrium (SUE) condition as follows: no user can reduce his or her perceived travel time by unilaterally changing routes. The team has extended this static SUE condition to the dynamic context and defined a time-dependent mode choice SUE (TDMSUE) as follows: For each O-D pair w and for each assignment–departure time interval t, no traveler can reduce his or her perceived mode travel cost or disutility by unilaterally changing modes.

Given the assumptions and definitions above, the mode choice problem is to find a SUE mode flow pattern, $y = \{y_m^{wt} | \forall w \in W, t \in T, m \in M\}$, satisfying the TDMSUE definition. This essentially means that the attribute values used in the mode choice model, particularly the LOS attributes, are mutually consistent with the results of the traffic simulation assignment model obtained by splitting the travelers to modes and subsequently assigning them to routes.

Multiple Attributes (Criteria) in Highway Route-Choice Process

The solution of the mode choice problem gives an equilibrium mode flow pattern, y_m^{wt}, $\forall w \in W$, $t \in T$, $m \in M$, which forms the input for a multiclass dynamic user equilibrium traffic-assignment problem. In other words, the mode choice model provides the relative fractions of users by different modes, including those whose choices entail automobile use as driver or passenger on the highway network. The path choices of these users and the associated network performance measures are consistent with the resulting multiclass dynamic user equilibrium problem defined here. The main features of the problem addressed here entail the response of users not only to attributes of the travel time experienced on average by travelers on a particular path at a given time, but also to the prices or tolls encountered and the reliability of travel time.

Accordingly, users are assumed to choose a path that minimizes a generalized cost or disutility that includes three main path attributes: travel time, monetary cost, and a measure of variability to capture reliability of travel. Denote by $GC_k^{wtm}(\alpha)$ the experienced route generalized cost perceived by travelers with VOT α between O-D pair w departing at time t and using mode m and route k:

$$GC_k^{wtm}(\alpha, \beta) = TT_k^{wtm} + \alpha \times TC_k^{wtm} + \beta \times TTSD_k^{wtm}, \quad \forall k \in K(w, t, m) \qquad (4.2)$$

where TT and TC denote the travel time and travel cost, respectively, and $TTSD$ is the standard deviation of the travel time with subscripts and superscripts as already defined for route k, mode m, and O-D pair w. The parameters α and β denote the coefficients of the travel cost and travel time standard deviation, respectively; they are interpreted as VOT and value of reliability, respectively, as discussed in the previous chapters.

Recognizing Highway User Heterogeneity

In the above generalized cost expression, the parameters α and β represent individual trip maker's preferences in the valuation of the corresponding attributes. As shown in Chapter 3, these preferences vary across travelers in systematic ways that may be captured through user sociodemographic or trip-related attributes or in ways that may not be directly observable. Variation in user preferences for the different travel choice attributes is referred to as *heterogeneity.* Both α and β in the generalized cost expression above may vary across users. However, to realistically capture the effect of pricing and its impact on different user groups, it is essential to represent the variation of user preferences in response to prices, captured here through the parameter α. Accordingly, the focus is on capturing heterogeneous VOT preferences across the population of highway users. Preferences for reliability may also reflect heterogeneity, and the same approach used here for VOT may be extended to incorporate both. It should be noted, however, that the empirical estimation analysis reported in the previous chapter suggests that heterogeneity in VOT is much more significantly present than in the preferences for reliability.

To reflect highway user heterogeneity, VOT in this study is treated as a continuous random variable distributed across the population of travelers, with a density function $\phi(\alpha) > 0$, $\forall \alpha \in [\alpha^{min}, \alpha^{max}]$, and $\int_{\alpha^{min}}^{\alpha^{max}} \phi(\alpha) d\alpha = 1$, where the feasible range of VOT is determined by a given closed interval $[\alpha^{min}, \alpha^{max}]$. The distribution of VOT can be calibrated using discrete choice modeling techniques with random coefficient specifications, as shown in Chapter 3 using New York data. Accordingly, the VOT distribution is assumed known for assignment purposes. As discussed, in the New York application the corresponding coefficient β is taken as a constant across users when α is specified as a random variable distributed across the user population.

The time-dependent O-D demands for the entire range of VOT over the planning horizon (i.e., $y^{wtm}(\alpha)$, $\forall w \in W$, $t \in T$, $m \in M$, $\alpha \in \lfloor \alpha^{min}, \alpha^{max} \rfloor$) can be obtained based on a given mode flow pattern y_m^{wt}, $\forall w \in W$, $t \in T$, $m \in M$.

The analyst is interested in solving for $x_k^{wtm}(\alpha)$, $\forall w \in W$, $t \in T$, $m \in M$, $k \in K(w, t, m)$, $\alpha \in \lfloor \alpha^{min}, \alpha^{max} \rfloor$, the route flow for users with VOT α between O-D pair w departing at time t and using mode m and route k, and the corresponding experienced route generalized cost $GC_k^{wtm}(\alpha)$.

Note that although the mode choice utility function also reflects one's VOT, and in some instances reliability, the specifications used in practice for mode choice and route choice tend to differ markedly, and hence may not be directly comparable. For instance, mode choice models incorporate a richer array of sociodemographic attributes, as well as more refined definitions of travel time components (e.g., waiting time versus in-transit time) that affect mode valuation differentially. Accordingly, it is not expected that coefficients will be the same in models for these different choice dimensions. Ideally, a more complete activity-choice model formulation may achieve consistency in attribute valuation, but this is not expected in the models currently used in practice.

Generating Reliability Measures

As noted, travel time reliability is another important measure in the choice procedure. To incorporate this measure in the integrated model, it is necessary to devise a method to generate it for the respective paths and O-D pairs in connection with the movement of vehicles through the network. The reliability measure in this application is generated by a method that exploits a robust relation found to hold between the standard deviation and the mean values of the travel time per unit

distance. This method is best applied directly at the path level (between given origin and destination). In addition to the relation proving more robust at the path (O-D) level than at the link level, this approach circumvents the need to include this attribute in shortest-path labeling procedures. It also circumvents the challenging issue of link travel time correlations and the resulting nonadditivity across links defining a given path.

It is observed that, in general, the standard deviation of travel time per distance increases with its mean value. It is assumed that there is a linear relationship between the mean travel time per unit distance and its standard deviation, which is

$$TTSDMILE = b_0 + b_1 \times MEANTTMIL \tag{4.3}$$

The method leverages a robust relation that is shown to hold between the standard deviations and the mean values of the travel time per unit distance. First established in research conducted by Mahmassani and co-workers (Mahmassani and Herman 1987; Mahmassani and Tong 1986; Chang and Mahmassani 1988; Stephan and Mahmassani 1988) using actual traffic observations, the relation has been investigated in greater depth using both actual traffic data and simulation experiments. This relation has been shown to be more robust at the path (O-D) level than at the link level. By using this relation, the proposed solution algorithm circumvents the need to include link-level reliability measures in shortest-path labeling procedures and the challenging issue of correlations of link travel time.

In this study, the relation between standard deviation (reliability measure) and the mean values of the travel time per unit distance is estimated using a linear regression model based on simulated vehicle trajectories on the network (here, the New York metropolitan regional network) under consideration. From the regression analysis (see Figure 4.1), the reliability measure is generated using the equation represented in the figure.

This approach provides a very efficient procedure for combining reliability sensitivity with heterogeneous preferences

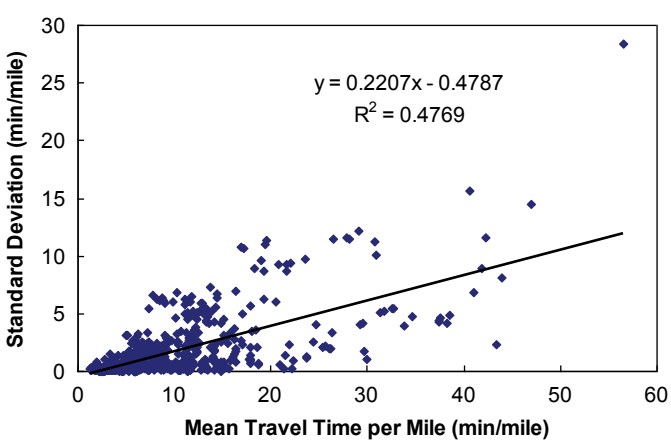

Figure 4.1. Standard deviation versus mean of travel time per unit distance.

for mean and standard deviation of travel time using the generalized cost function in Equation 4.2.

Multicriterion Dynamic User Equilibrium Route Choice Decision

A key behavioral assumption made for the route choice decision is that each trip maker would choose a route that minimizes the route generalized cost function. Specifically, for trips with VOT α, a path $k^* \in K(w, t, m)$ will be selected if and only if $GC_{k^*}^{wtm}(\alpha) = \min\{GC_k^{wtm}(\alpha) | k \in K(w, t, m)\}$. Based on this assumption, the multiclass multicriterion dynamic user equilibrium (MDUE), a multiclass multicriterion and dynamic extension of Wardrop's first principle, is defined as follows: For each O-D pair w, mode m, and assignment–departure time interval t, no traveler can reduce his or her experienced route generalized cost with respect to his or her particular VOT α by unilaterally changing path.

This definition implies that, at MDUE, each traveler is assigned to a path with the least generalized cost with respect to his or her own VOT. The next section presents an integrated multidimensional network choice model framework for the problem under consideration.

Conceptual Framework: Integrated Multidimensional Network Choice Model

To solve the integrated dynamic traveler mode–path assignment problem in multimodal transportation networks, the analyst essentially wants to determine the number of travelers for each alternative (i.e., mode and path) and the resulting temporal–spatial loading of vehicles and travelers. The sequence of the integrated multidimensional network choice is traveler's mode choice, ride-sharing choice, and vehicle generation, as well as vehicle route choice and simulation.

Figure 4.2 shows the integrated multidimensional simulation-based dynamic microassignment conceptual framework. This framework gives the procedure and evolution of traveler's equilibrium mode choice, ride-sharing choice, vehicle generation, vehicle equilibrium route choice, and traffic simulation.

Time-Dependent Traveler Origin–Destination Demand and Characteristics

Demand input for the integrated model consists of a set of time-dependent traveler O-D trips and corresponding individual characteristics, such as income, auto ownership, purpose, and VOT, which are used in both mode and route choice procedures. In this study, the time-dependent traveler O-D demand is generated directly from the New York Metropolitan Transportation Council's best practice model (NYBPM)

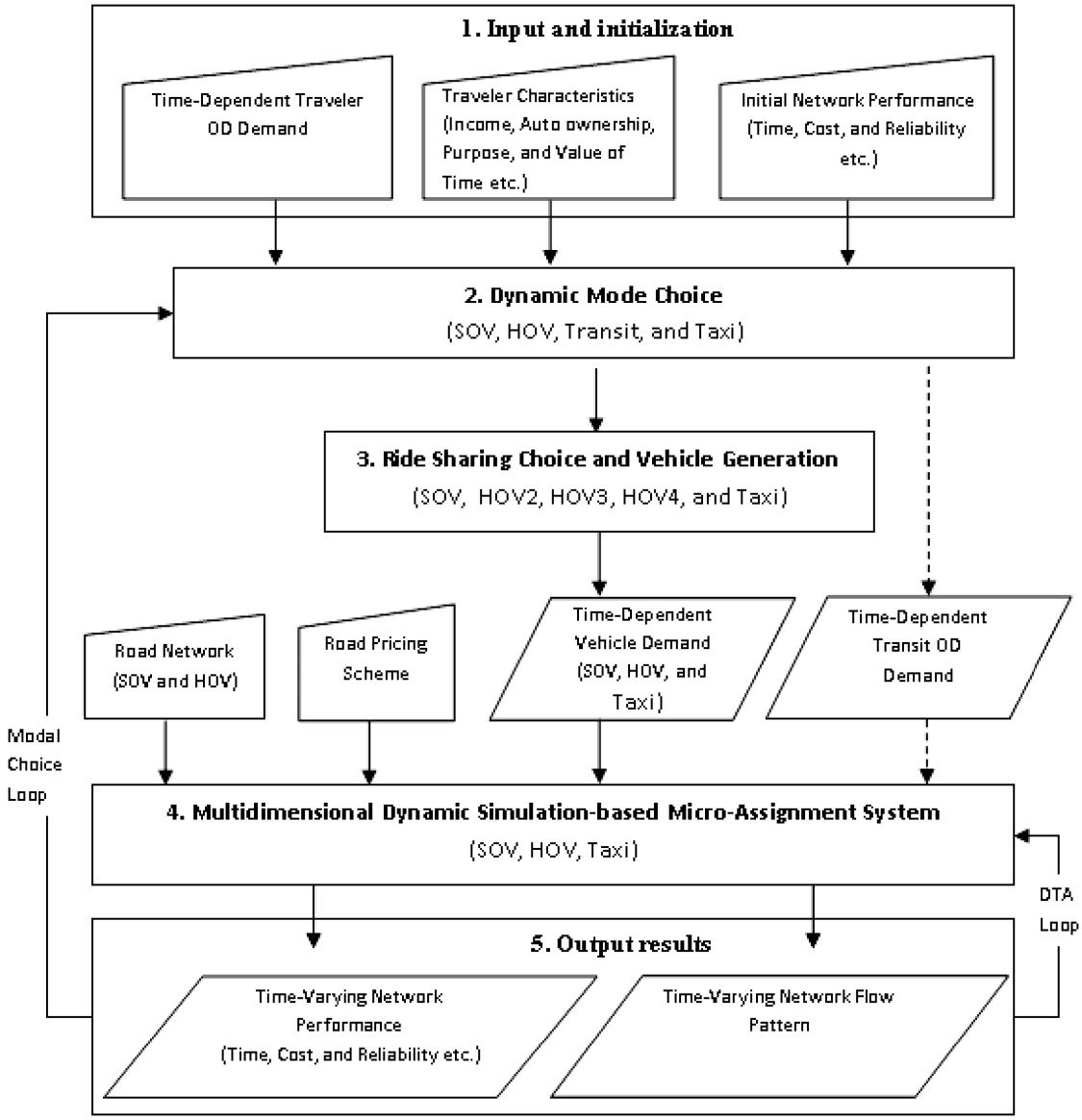

Figure 4.2. Framework of the integrated multidimensional network choice model.

(Vovsha, Donnelly, and Gupta 2008; Vovsha, Donnelly, and Chiao 2008; Chiao and Vovsha 2006; PB Consult, Inc. 2005; Vovsha et al. 2002). The NYBPM is a static tour-based model that includes individual characteristics by a microsimulation model based on social and economic characteristics of the population, such as employment status. From a set of given time-of-day (TOD) distributions in the NYBPM tour-based model, 30-minute interval time-varying traveler O-D trips with 13 modes can be generated for interval traveler trips between internal zones (i.e., zones inside the NYC regional network). External demand comprises trips from, to, or between external zones (i.e., zones outside the NYC regional network); these trips are all auto trips. External vehicle trips are not involved in the mode choice part, but will be considered in the route choice and network simulation part. In addition to the TOD pattern generated from the NYBPM model, a 15-minute interval auto trip (i.e., single-occupant vehicle [SOV] and high-occupancy vehicle [HOV]) TOD pattern is estimated based on historical detector data. This study combines the TOD patterns from both models to define the final departure-time pattern.

Accordingly, demand in this study consists of individual travelers, and the mode choice is a disaggregated choice or microassignment procedure; that is, each traveler selects his or her best mode based on a Monte Carlo simulation technology and choice probabilities of available alternatives in his or her choice set. Note that this framework can further integrate an activity chain–based DTA model by replacing the underlying trip-based DTA model with an activity-based DTA model. In this study, the integrated model is restricted to a trip-based model.

Nested Logit-Based Mode Choice Model

With the above time-dependent traveler O-D demand with individual characteristics for internal zones, the mode choice model is used to determine the mode to be chosen for each traveler according to his or her individual characteristics (e.g., income, auto ownership, age) and mode attributes (LOS) at each departure time.

There are 13 modes considered in this study, as shown in the nested logit mode choice model structure in Figure 4.3. Note that the auto travelers (i.e., SOV, HOV2, and HOV3) can further choose their corresponding routes with minimum experienced generalized cost with respect to their VOT in the MDUE route choice and network simulation procedure. More importantly, this study allows flexible forms of both mode and route choice models; that is, it does not restrict the cost function of route choice function to be the same as the mode choice model. This flexibility is essential in practice, as specifications and calibrations of mode choice and route choice model may use different functions and data. Therefore, both mode choice and route choice models can be used with any well-calibrated choice models and utility functions.

The mode choice model in this study is a nested logit model, which is appropriate when the set of alternatives faced by a decision maker can be partitioned in subsets (nests). In multimodal regional transportation systems, travelers face drive-alone, ride-sharing, transit, and taxi choices. The nested logit mode choice model shown here reflects one of the best combinations of individual characteristics and mode attributes in the utility functions. The utility function is shown in Equation 4.4:

$$U_m^{wt} = V_m^{wt} + V_{Bi}^{wt} + V_{Bj}^{wt} + \varepsilon_{m|j|i}^{wt} + \varepsilon_{Bi}^{wt} + \varepsilon_{Bj|i}^{wt} \tag{4.4}$$

where

V_m^{wt} = systematic utility for mode m;
V_{Bi}^{wt} = systematic utility for nest B_i, $\forall i = 1,2,3$ in Level 1;
V_{Bj}^{wt} = systematic utility for nest B_j, $\forall j = 1,2,\ldots,7,8$ in Level 2;
$\varepsilon_{Bi}^{wt}, \varepsilon_{Bj|i}^{wt}, \varepsilon_{m|j|i}^{wt}$ = random terms for each level that are independent to each other and are identically and independently distributed extreme value with scale parameters μ_1, μ_2, μ_3, respectively;
VL_{Bi} = logsum for nest B_i; and
VL_{Bj} = logsum for nest B_j.

Accordingly, the nested logit choice probabilities, marginal probabilities, and conditional probabilities can be evaluated by the following equations:

$$\Pr(i) = \frac{e^{VL_{Bi}(i)\mu_1}}{\sum_{i'=1}^{3} e^{VL_{Bi}(i')\mu_1}} = \frac{1}{\sum_{i'=1}^{3} e^{[VL_{Bi}(i') - VL_{Bi}(i)]\mu_1}} \tag{4.5}$$

$$\Pr(j|i) = \frac{e^{VL_{Bj}(j)\mu_2}}{\sum_{j' \in B_i} e^{VL_{Bj}(j')\mu_2}} = \frac{1}{\sum_{j' \in B_i} e^{[VL_{Bj}(j') - VL_{Bj}(j)]\mu_2}} \tag{4.6}$$

$$\Pr(m|j|i) = \frac{e^{V(m)\mu_3}}{\sum_{m' \in B_j} e^{V(m')\mu_3}} = \frac{1}{\sum_{m' \in B_j} e^{[V(m') - V(m)]\mu_3}} \tag{4.7}$$

$$\Pr(m) = \Pr(m|j|i) \times \Pr(j|i) \times \Pr(i) \tag{4.8}$$

$$p_m^{wt} = \Pr(m) \tag{4.9}$$

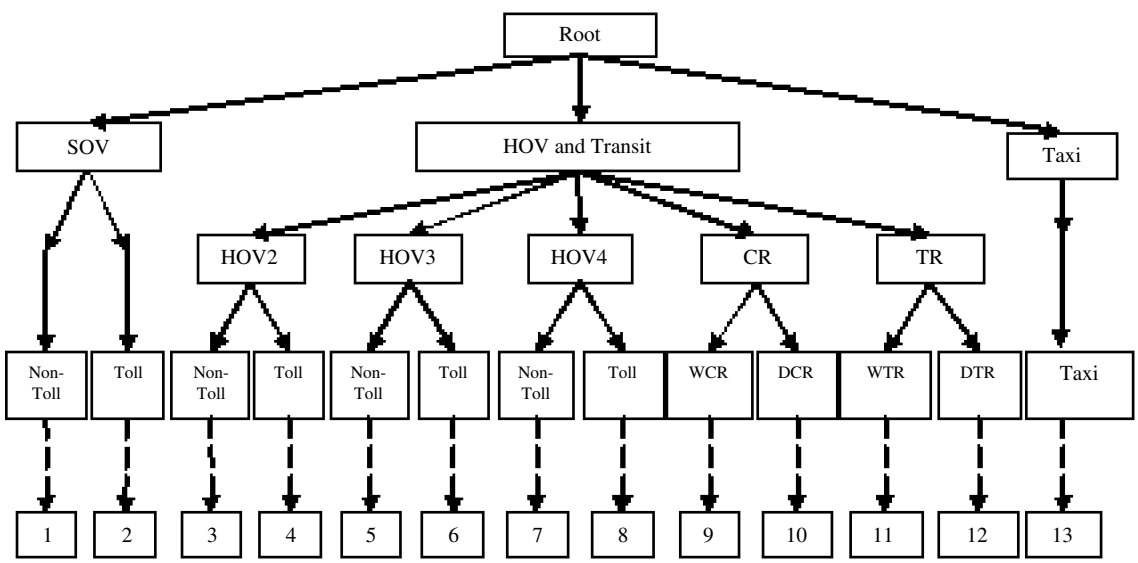

Figure 4.3. Nested logit mode choice model.

The result of the mode choice model is to assign a mode to each traveler. Because the majority of traffic interaction in the transportation networks is vehicle to vehicle, especially for highway networks, it is necessary to map travelers to vehicles, especially for those ride-sharing travelers, according to occupancy levels. The next section gives a ride-sharing choice and vehicle-generation procedure to connect the mode choice results to vehicle trips, which is the input for the MDUE route choice and network simulation model.

Ride-Sharing Choice and Vehicle-Generation Model

The mapping from travelers with mode choice to vehicle trips is a key element in the integrated model because it provides the essential demand input to the route choice and dynamic network simulation and assignment procedure. Most available regional models distinguish trips between internal zones and trips from, to, or between external zones; only internal trips are involved in the mode choice procedure, while modes of external trips are predefined. However, in the network simulation and assignment procedure, internal vehicle trips and external vehicle trips will be interacted with each other. Accordingly, the proposed ride-sharing choice and vehicle-generation model includes two components: ride-sharing choice and internal vehicle generation, and appending external vehicle trips to the vehicle demand.

The ride-sharing choice is intended to address the carpooling behavior of HOV travelers, which is a procedure to map travelers to vehicles based on origin, destination, and departure time, as well as occupancy level (i.e., HOV2, HOV3, and HOV4). For a set of travelers with the same origin, destination, departure time, and occupancy level, there are three mapping methods: (1) deterministic simple mapping, (2) deterministic sorted mapping, and (3) random mapping. Deterministic simple mapping simply selects *s* (where *s* is the occupancy level) travelers in the traveler set sequentially. Deterministic sorted mapping selects *s* travelers sequentially in a sorted traveler set in which the sorted criterion can be the VOT of each traveler. Random mapping uses a Monte Carlo simulation process to randomly select *s* travelers in the traveler set. Following the procedure of ride-sharing choice of HOV travelers, this procedure generates internal vehicles based on the results of ride-sharing choice. After generating vehicles for all the internal zones, this model appends vehicles for external zones. The procedure of ride-sharing choice and vehicle generation is shown in Figure 4.4.

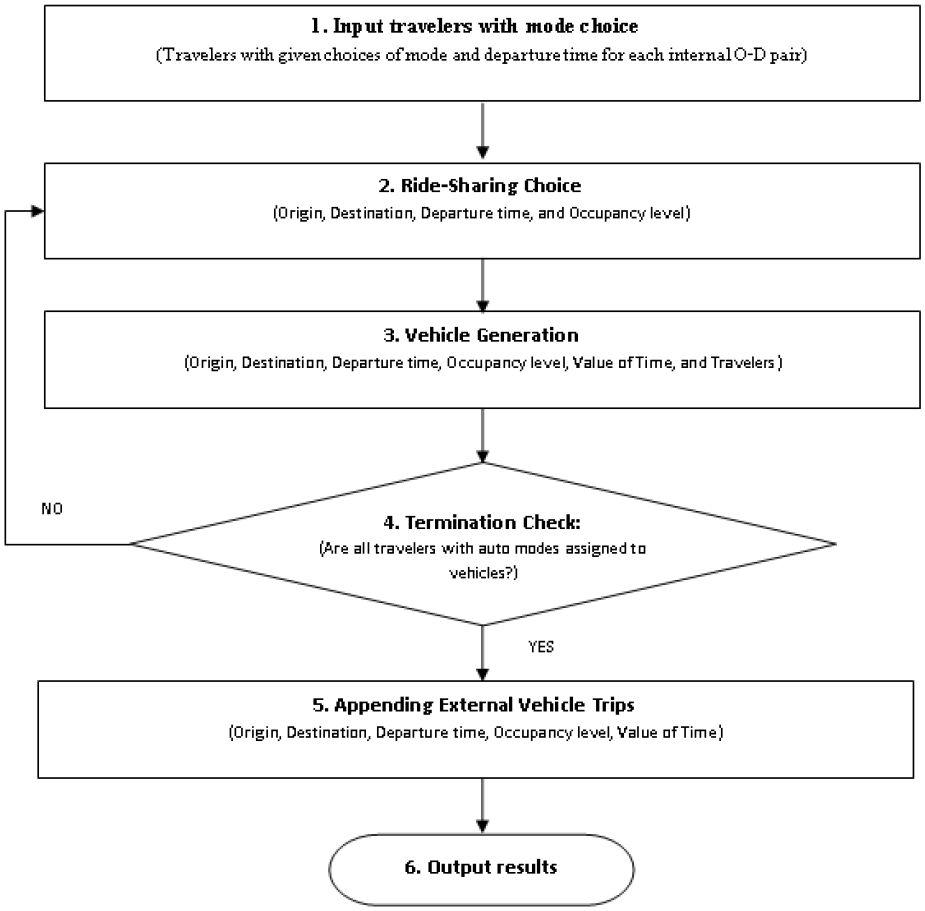

Figure 4.4. Ride-sharing choice and vehicle-generation model.

The mode choice model, ride-sharing choice, and vehicle-generation procedures result in a time-varying vehicle O-D demand pattern that consists of a set of multiclass vehicles with distinct individual origin, destination, departure time, occupancy level, VOT, and so forth, for the route choice and network simulation procedure. The next section describes the multidimensional simulation-based dynamic microassignment system used in the route choice and network simulation procedure in this study to assign routes to each vehicle.

Multidimensional Simulation-Based Dynamic Microassignment System

To support pricing and congestion, a multidimensional simulation-based dynamic microassignment system was developed to address the MDUE route choice behavior. The system features the following three components: (1) traffic simulation (or supply), (2) traveler route choice behavior, and (3) path set generation. These components have become relatively standard in state-of-the-art simulation-based assignment procedures, following a blueprint originally developed for FHWA in the form of the DYNASMART simulation–assignment methodology (Jayakrishnan et al. 1994). Accordingly, the simulation capabilities used in this work are interchangeable with almost any particle-based simulator that tracks individual vehicle trajectories in a micro or meso flow-modeling framework. However, the algorithmic procedures for equilibrium seeking differ across platforms. In particular, the algorithms for finding a TDUE with heterogeneous users have, to the team's knowledge, only been implemented in conjunction with DYNASMART-P. However, these procedures could be adapted with most of the microassignment tools. The traffic simulator—DYNASMART (Jayakrishnan et al. 1994) in this case—is used to capture the traffic flow propagation in the traffic network and evaluate network performance under a given set of mode and route decisions made by the individual travelers. Given user behavior parameters, the traveler route choice behavior component aims to describe travelers' route selection decisions (i.e., the MDUE route choice model in this study). The third component, path set generation, is intended to generate realistic route choice sets for solving the traveler assignment problem. Figure 4.5 depicts the flowchart of the system.

A general overview and recommended methods for a simulation-based solution approach to the TDUE assignment problem are presented below, and mathematical formulations and solution algorithms are presented in the Mathematical Formulations of the Integrated Multidimensional Network Choice Model and Solution Algorithms for the Integrated

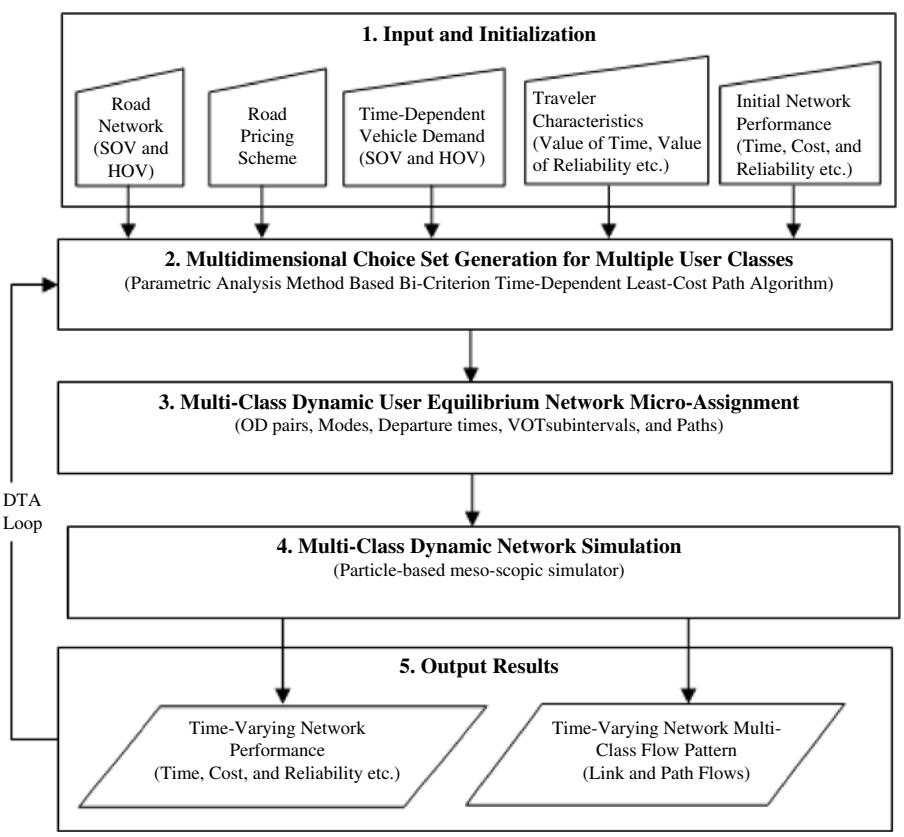

Figure 4.5. Multidimensional simulation-based dynamic microassignment system.

Multidimensional Network Choice Model sections within Appendix A. The following sections present an overview of a simulation-based iterative solution framework to solve the integrated model and present the some key issues in the integrated model, specifically route choice and path computations including reliability, a column-generation solution framework, and algorithms and challenges for large-scale applications.

Simulation-Based Iterative Solution Framework

Overview of Solution Framework

The TDMSUE-MDUE problem is to find both equilibrium travelers' mode choice and equilibrium vehicles' route choice with a given time-dependent traveler demand. The TDMSUE problem is solved by a projected gradient-based descent direction method. However, it is not practical to enumerate the complete set of feasible routes for solving the MDUE problem in a realistically sized transportation network. To capture the individual choice behavior and traffic dynamics, the simulation-based DTA algorithmic framework disaggregates the O-D demands into individual vehicles. Only a portion of paths would have a nonzero probability of carrying vehicles in an MDUE solution. This study uses the trajectories of vehicles as a proxy to store the feasible path set, using what is referred to as the vehicle-based implementation technique (Lu et al. 2007), to optimize computer memory use and eliminate many otherwise unrealistic paths. To avoid explicit enumeration of all feasible routes, the study applies a column-generation approach to generate a representative subset of paths with competitive costs to augment the feasible path set. The parametric analysis method (PAM) is applied to obtain a set of breakpoints that partition the entire VOT interval into multiple subintervals. A projected descent direction method is used to solve the resulting MDUE problem in a restricted (reduced) path set; this is called the restricted multiclass multicriterion dynamic user equilibrium (RMDUE) problem in equations (see Solution Algorithm for the Integrated Multidimensional Network Choice Model section in Appendix A for more detail).

Route Choice and Path Computations Including Reliability

The main impediment for solving the MDUE problem of interest is due largely to the relaxation of VOT from a constant to a continuous random variable. This relaxation leads to the need to find an equilibrium state resulting from the interactions of (possibly infinitely) many classes of trips, each of which corresponds to a class-specific VOT, in a network. If, in the extreme case, each trip maker (or class) requires its own set of time-dependent least-generalized cost paths, finding and storing such a grand path set is computationally intractable and memory intensive in (road) network applications of practical sizes. In order to circumvent the difficulty of finding and storing the least-generalized cost path for each individual trip maker with different VOT, PAM is proposed to find the set of extreme efficient path trees, each of which minimizes the parametric path generalized cost function (Equation 4.2) for a particular VOT subinterval. The idea of finding the set of extreme efficient paths to which heterogeneous trips are to be assigned is based on the assumption (Dial 1997; Marcotte and Zhu 1997) that in the disutility minimization-based path choice modeling framework with convex disutility functions, all trips would choose only among the set of extreme efficient paths corresponding to the extreme points on the efficient frontier in the criterion space.

Essentially, the PAM bicriterion time-dependent least-cost path (BTDLCP) algorithm has two important roles to play in solving the proposed problem: (1) it transforms a continuous distributed VOT into multiple user classes, and (2) it generates a time-dependent least-cost path tree for each user class, which subsequently defines a descent search direction for the MBDUE traffic assignment problem.

A novel approach is developed for this study to generate and incorporate reliability measures in route choice and supporting path computation. The method leverages a powerful relation that is shown to hold between the standard deviation and the mean values of the travel time per unit distance. By combining this approach with the multiple paths produced by the parametric shortest-path methods used to reflect heterogeneity in user preferences, a very efficient implementation has been devised and tested in this work. Thus, for the given path set corresponding to the various classes of users (determined by the parametric shortest-path procedure), the corresponding reliability measure is estimated directly at the path level, and relabeling of the paths is then performed for the various classes, taking reliability valuation into consideration. This approach provides a very efficient procedure to combine reliability sensitivity with heterogeneous preferences for mean and standard deviation of travel time consistent with the generalized cost expression in Equation 4.2.

The PAM-based path-generation procedure including the reliability measure is shown in Figure 4.6.

Implementation Steps of Solution Framework

The simulation-based column-generation iterative solution framework for the TDMSUE-MBDUE problem includes four main steps, depicted in Figure 4.7. The steps are

1. Input and initialization;
2. Nested logit mode choice;
3. Ride-sharing choice and vehicle generation; and

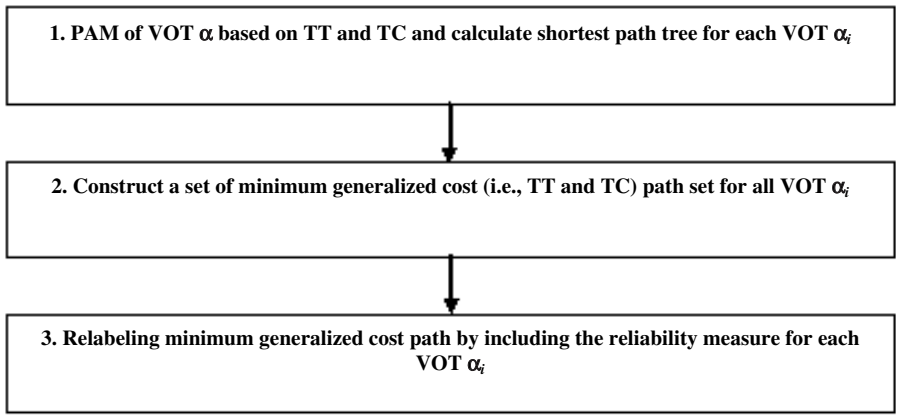

Figure 4.6. PAM-based path-generation procedure including reliability measure.

4. Multidimensional simulation-based dynamic micro-assignment.

Column Generation–Based MDUE Solution Algorithm

The column generation–based MDUE solution algorithm incorporating different vehicle classes (low-occupancy vehicles [LOVs] and HOVs) is outlined below, and its flowchart is presented in Figure 4.8.

Algorithms and Challenges for Large-Scale Applications

Computation and storage of least-generalized cost paths for time-dependent equilibrium problems constitute the major computational challenge for the algorithms described in this section. The parametric shortest-path procedures for random coefficient utility models represent a major breakthrough that allows consideration of heterogeneous users in a practical network setting. To alleviate the memory-demanding requirements of the flow-based DTA models, a vehicle-based technique is implemented. Vehicle-based implementations circumvent storing the grand path set and path assignment sets explicitly whereby the path information is extracted from vehicle trajectories, and thus provide considerable savings in memory requirements in the process. To this end, vehicle-based implementations of equilibrium methods and parametric shortest-path procedures are two major advances.

Nonetheless, the scale of the networks of interest imposes additional computational burdens on the solution algorithm that require further considerations of the design of the algorithms and their implementation schemes. These are developed and illustrated on the New York regional network in this study. The bottleneck here is again the computational time required for time-dependent least-generalized cost path calculation. Two features in the implementation of PAM are introduced in conjunction with the New York application: (1) adjusting the step size in PAM and (2) gap-based shortest-path selection. Both features aim at reducing the computation time.

Step Size Adjustment in Parametric Analysis Method

In the outer loop, PAM is invoked to find the set of bicriterion time-dependent extreme efficient paths, to which all the trips with different VOTs are assigned, and the corresponding set of breakpoints (i.e., VOTs, $\boldsymbol{\alpha}_\mathbf{I} = \{\alpha_0, \alpha_1, \ldots, \alpha_i, \ldots, \alpha_I | \alpha^{\min} = \alpha_0 < \alpha_1 < \cdots < \alpha_i < \cdots < \alpha_I = \alpha^{\max}\}$) that partitions the entire feasible range of VOT $[\alpha^{\min}, \alpha^{\max}]$ and hence defines multiple classes of trips, where each class includes the trips with VOT $\alpha \in [\alpha_{i-1}, \alpha_i)$, $\forall i = 1, \ldots, I$. Starting from the lowest possible VOT, the bicriterion time-dependent shortest-path algorithm continuously solves for the time-dependent least-generalized cost (TDLGC) path tree rooted at each destination for a given VOT interval and determines the upper bound of that VOT interval, for which the TDLGC path tree remains optimal, until reaching the highest possible VOT. In order to move from the current VOT segment to the next one and obtain a different TDLGC path tree, a small value Δ needs to be added to the current breakpoint α_i. This implies that travelers cannot distinguish differences in VOT below Δ per minute. The value of Δ also implicitly sets an upper bound for the number of VOT segments generated in PAM, with a value of $\frac{(\alpha^{\max} - \alpha^{\min})}{\Delta}$. The feasible range of VOT is given by the closed interval $[\alpha^{\min}, \alpha^{\max}]$ and can be estimated from survey data. As a result, Δ is a fixed given value. PAM requires a full run of BTDLCP calculations for finding one VOT segment, which is time consuming on a large-scale network. In order to reduce the computational time, Δ can be set to a larger value. As indicated elsewhere, Δ implies the indifference band in VOT

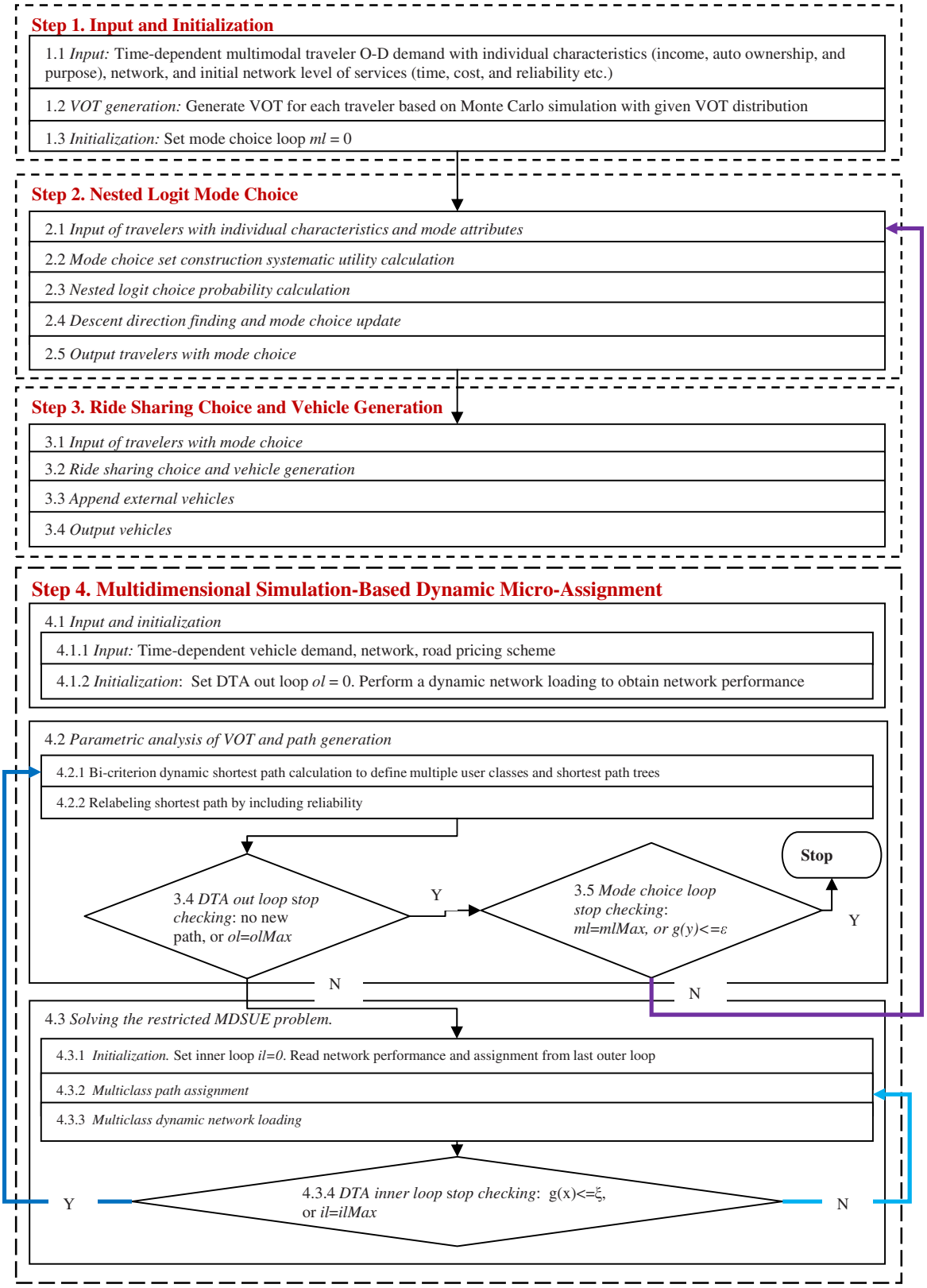

Figure 4.7. Simulation-based column-generation solution framework.

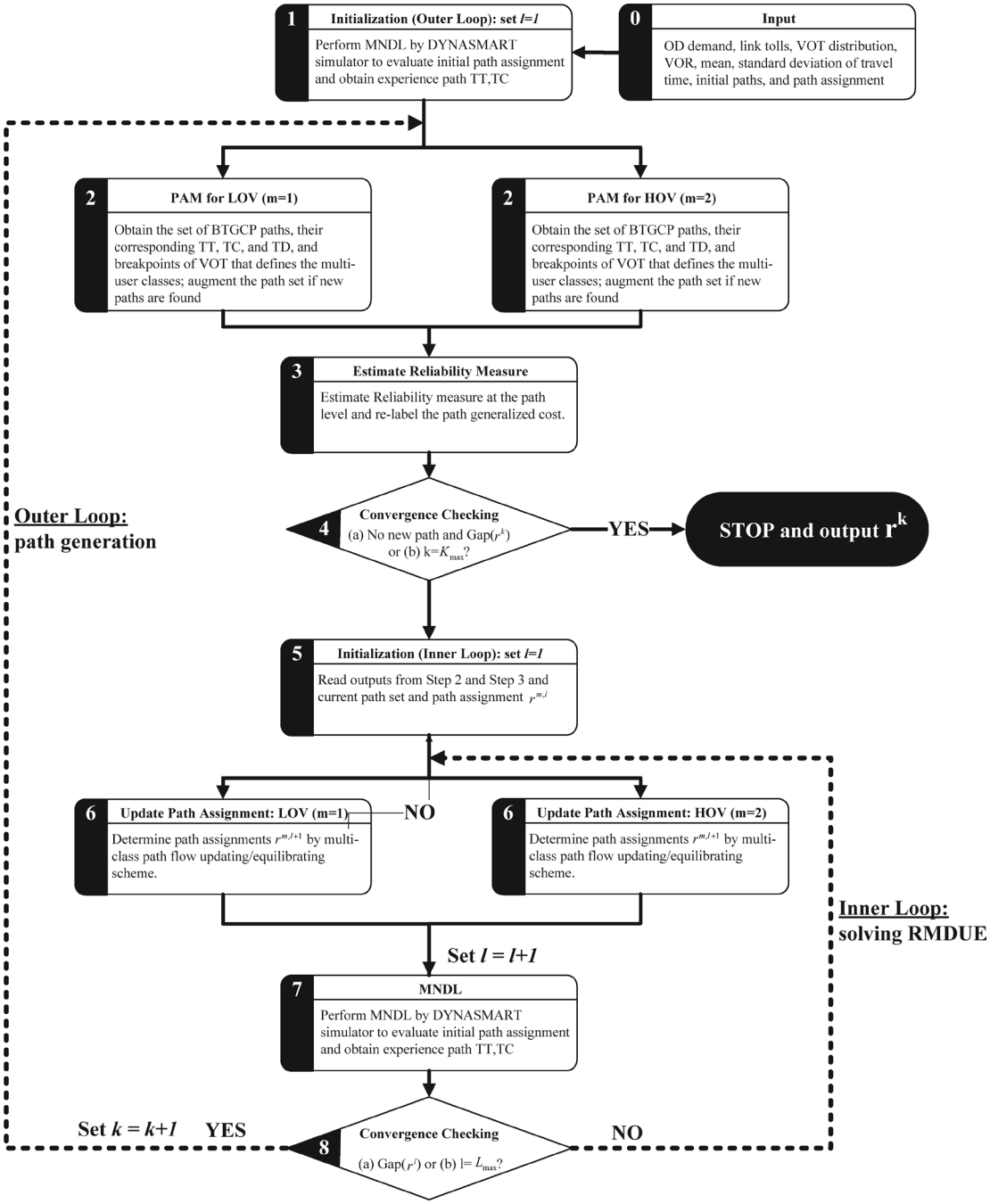

Figure 4.8. Flowchart of MDUE solution algorithm.

among travelers and should be set to a small value. Increasing the value of Δ in a small network may lead to inaccurate and unrealistic predictions of flow distribution patterns, whereas in a large-scale network the flow patterns will remain valid.

This claim can be validated through a simple example. For the sake of simplicity, the generalized cost does not include reliability in this example. However, for a given path, reliability does not vary with VOT; therefore, the following illustration remains valid once the reliability term is added into the generalized cost. In PAM, the generalized cost perceived by travelers with VOT α from O-D pair w at departure time t along path $k \in \tilde{K}(w, t, m, \alpha_i)$ is given in Equation 4.2.

In PAM, the BTDLCP algorithm calculates a time-dependent least-cost path tree for a given α rooted at every destination node based on the generalized cost described in Equation 4.2. The scale of the network influences the algorithm with regard to the proportion of the paths using tolled links in the path set. In this extreme case, every path found by BTDLCP uses the tolled link, thus making the algorithm very sensitive to VOT. If the value of α is changed to $\alpha + \Delta$, another path tree may

be obtained that differs considerably from the current one in a small network. When Δ is large, a lot of information may be lost when generating the least-cost path tree, resulting in inaccurate flow distribution patterns for a small network. This inaccurate result will be avoided in large networks due to only a small portion of paths using tolled links; thus, the least-cost path trees are more robust. In this case, setting a larger value for Δ would not result in a significant loss in the total path set generated by the BTDLCP algorithm.

Gap-Based Selection Technique in Parametric Analysis Method

The other modification to reduce computational effort is a gap-based selection technique for time-dependent shortest-path calculation. Essentially, the path-finding algorithm is applied only to a fraction of the destination zones, selected on the basis of the gap values of vehicles arriving at that destination. After sorting the destination zones according to their gap values, PAM is invoked in the outer loop for the worst $1/n$ of these destination zones to obtain new VOT partitions and update the previous path set with new paths (if any) found by the bicriterion time-dependent shortest-path algorithm. For the rest of the destination zones, the path sets and VOT partitions remain the same.

An integrated multidimensional network modeling procedure for supporting congestion and pricing studies was presented above. In the next section, an application of the proposed integrated procedure using the NYBPM regional network is demonstrated.

Demonstration Using New York Regional Network

Building a Large-Scale Network Model: Summary of Challenges

In general, because of their ability to represent network operational characteristics, simulation-based DTA models require more detailed network information than comparable static assignment models. Traffic control signs and signals, left turns, and other movement capabilities at a node are mainly (and crudely) represented in the link performance function in a static network, whereas a dynamic model requires more accurate information on junction control and allowed movements at each phase at a signalized intersection, as well as careful definition of each downstream movement at a node.

Basically, there is no direct method of (correctly) converting a static network model into a dynamic network model in one shot using only the existing data obtained from the provided static network model database. Smart conversion of the existing database, use of external information sources, and more importantly, use of engineering judgment are essential parts of building a large-scale dynamic network model.

In sum, models developed for static assignment application generally exhibit a variety of drawbacks that render them inappropriate for dynamic network analysis, including

- Oversimplified representation of junctions, especially freeway interchanges, for correct operational simulation;
- Absence or incorrect control information at junctions and lack of a reliable electronic database of control devices and control parameters at signalized junctions;
- Poor definition of origin and destination zones, including treatment and connection of centroids and external traffic generators;
- Insufficient specification of the operational attributes of links and junctions for the purpose of traffic simulation; and
- Absence of time-varying O-D information, which must be synthesized from available static matrices, coupled with traffic counts sometimes taken in mutually different time periods.

Conversion of Existing Network for Dynamic Analysis

The regional NYBPM includes 28 counties from a tristate area divided into 3,586 internal traffic analysis zones. The counties include

- Ten counties from the New York Metropolitan Transportation Council area;
- Two other counties from New York State;
- Thirteen counties from the North Jersey Transportation Planning Authority area;
- One other county from the state of New Jersey; and
- Two counties from the state of Connecticut.

The zones are mainly concentrated in NYC; these include

- 2,449 zones from New York State;
- 740 zones from the state of New Jersey;
- 397 zones from the state of Connecticut; and
- 111 external zones for travel entries to and exits from the network.

The NYBPM network also contains

- 53,395 links; and
- 31,812 nodes.

The DTA model converted from the static TransCAD model can be seen in Figure 4.9, and Figure 4.10 shows the DYNASMART-P model.

Details of the conversion steps and procedures applied in the process are described in the Application to New York Regional Network section in Appendix A. These include procedures to

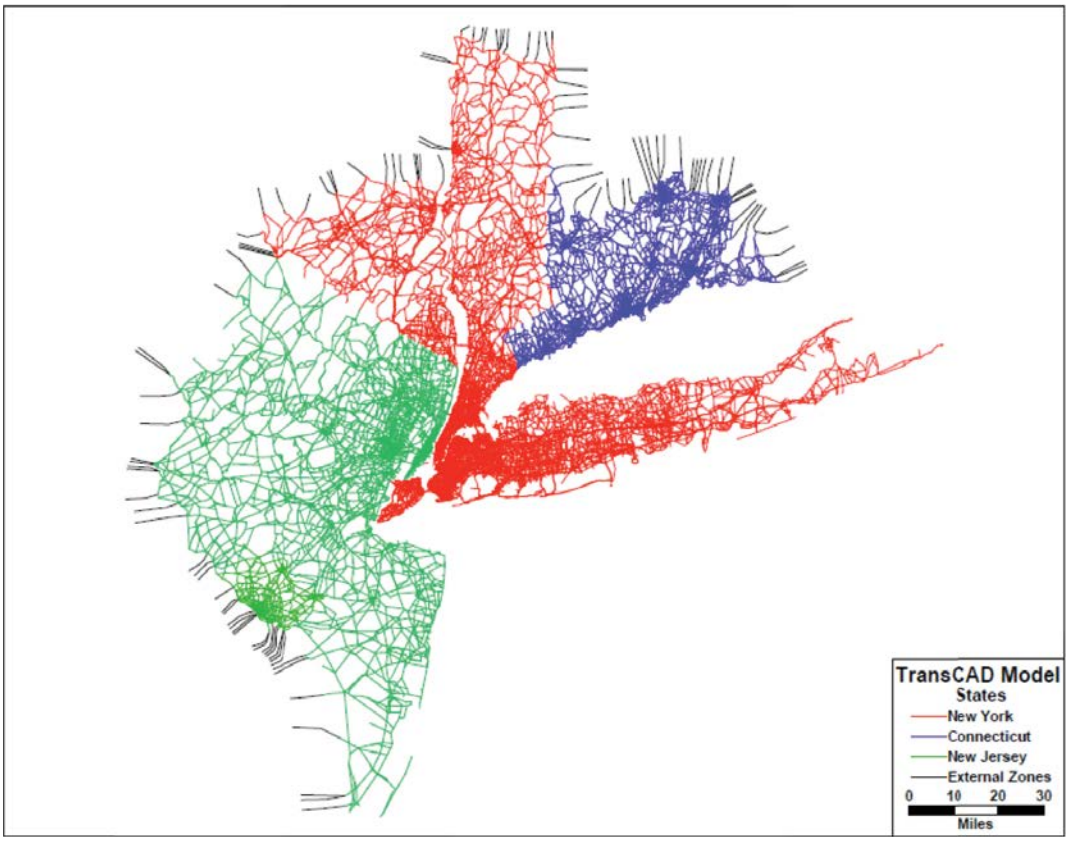

Figure 4.9. TransCAD model of the NYBPM network.

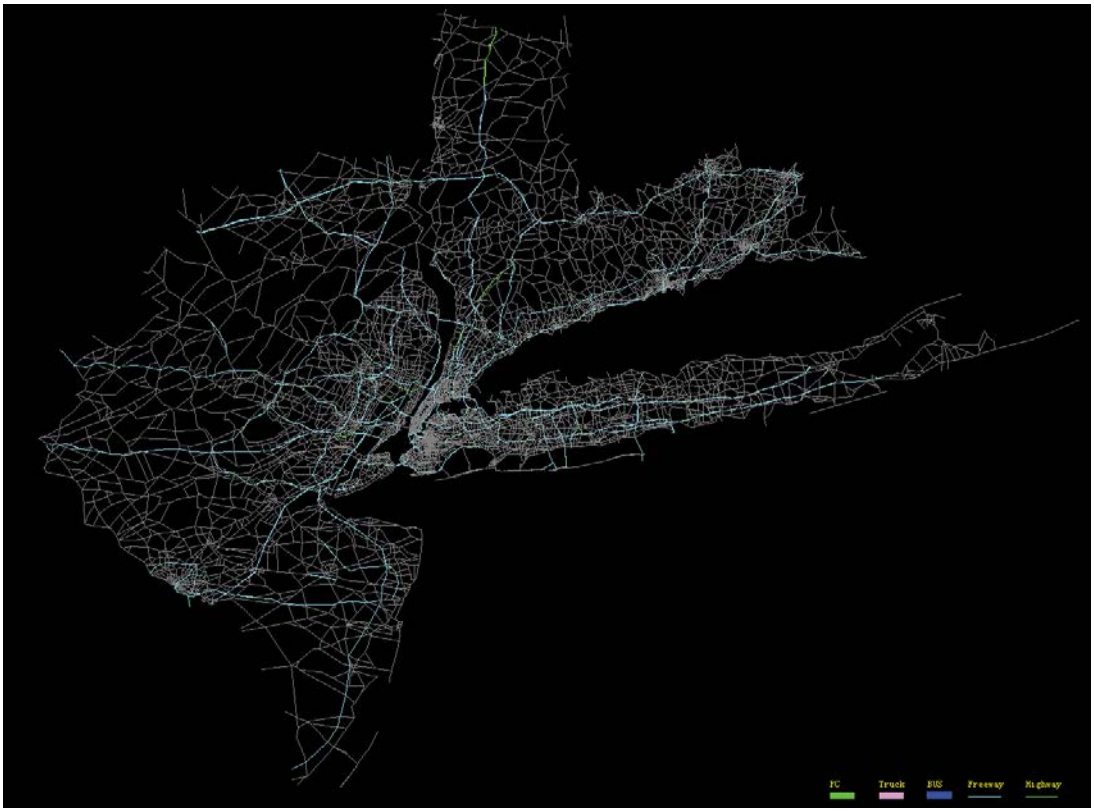

Figure 4.10. DYNASMART-P model of the New York network.

adjust the representation of geometric features of interchanges to support operational simulation and to assess, assign, and verify properties of junctions and specification of movements at junctions. The conversion also includes the preparation of the various required and optional input data files for the simulation. The properties of the resulting DYNASMART-P network are described below.

Zone Information

There are 3,697 zones. Of these, 3,586 are internal; only 111 are external.

Node and Control Information

There are 28,406 nodes. Control information for the nodes is as follows:

- 3,816 uncontrolled;
- 2,625 yield signed;
- 12,944 all-way stop signed;
- 8,054 actuated controlled; and
- 967 two-way stop signed.

Link and Type Information

The 68,490 links on the network are classified as follows:

- 6,026 freeways;
- 169 freeway HOV links;
- 56,102 arterials;
- 37 HOV arterial links;
- 150 highways;
- 2,688 on-ramps; and
- 3,318 off-ramps.

Pricing Information

There are 297 tolled links; of these, 291 use static tolling and only six use dynamic tolling. As seen in Figure 4.11, most of the pricing is nondistance based except along the I-95 New Jersey Turnpike corridor. Tolling on the major bridges and tunnels is as follows:

- The George Washington Bridge, Lincoln Tunnel, Holland Tunnel, Goethals Bridge, Outerbridge Crossing, and Bayonne Bridge are dynamically tolled bridges;

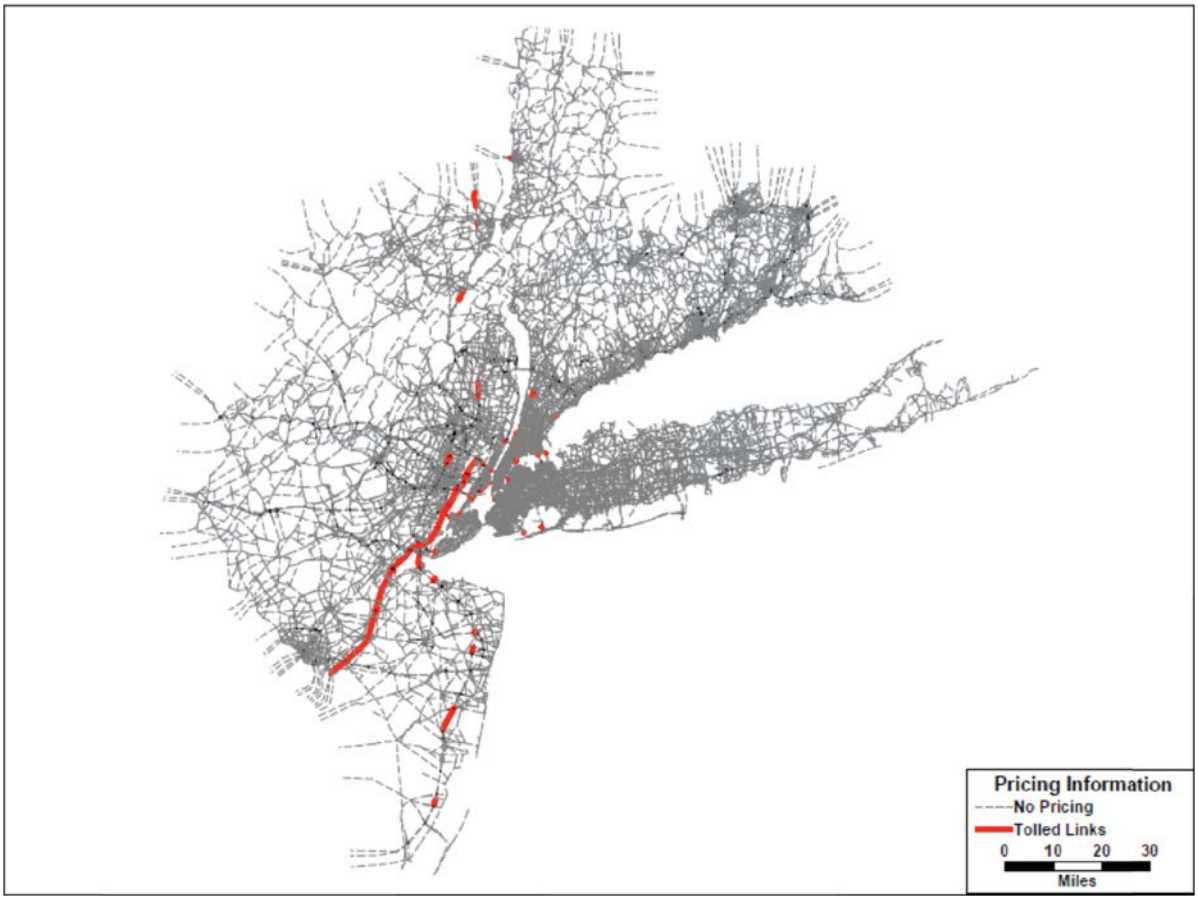

Figure 4.11. Pricing information for the New York network.

- The Verrazano-Narrows Bridge, Bronx-Whitestone Bridge, Brooklyn-Battery Tunnel, Queens Midtown Tunnel, Throgs Neck Bridge, Triborough Bridge, Marine Parkway-Gil Hodges Memorial Bridge, Cross Bay Veterans Memorial Bridge, and Henry Hudson Bridge are the bridges and tunnels tolled in New York metropolitan area; and
- The Tappan Zee Bridge, Bear Mountain Bridge, Kingston Rhinecliff Bridge, Mid Hudson Bridge, and Newburgh Beacon Bridge are the tolled bridges in New York State.

Methodology for Calibration of Origin–Destination Demand for Dynamic Analysis

Given static O-D demand information and time-dependent link measurements, the dynamic O-D demand estimation procedure aims to find a consistent time-dependent O-D demand table that minimizes the deviation between (1) estimated link flows and observed link counts and (2) estimated demand and target demand (based on the static demand matrix). The induced network flow pattern can be expressed in terms of path flows and link flows.

In a dynamic context, and especially in congested networks, elements of the mapping matrix between O-D demand and link flows are not constant and are, themselves, a function of the unknown O-D demand values. A bilevel dynamic O-D estimation formulation is adapted here in order to integrate the DTA constraint. Specifically, the upper-level problem seeks to estimate the dynamic O-D trip desires based on given link counts and flow proportions, subject to nonnegativity constraints for demand variables. The flow proportions are in turn generated from the dynamic traffic network loading problem at the lower level, which is solved by a DTA simulation program, with a dynamic O-D trip table calculated from the upper level.

The mathematical formulations and solution algorithms for the time-dependent O-D estimation process are detailed in Appendix A. Application of the procedures to the New York regional network are also described. Recognizing some of the data limitations described earlier, it was still possible to develop and calibrate a reliable DTA tool that represents the dynamics of traffic in the study area to a reasonable degree and allows meaningful comparative analysis of alternative scenarios. To evaluate the performance of the procedure, the root mean squared error between observed link volumes and simulated link volumes are used as an overall measure of effectiveness.

Validation against individual link counts was performed for selected links. Cumulative curves provide insight into the ability of the resulting assignment to capture the link flow volumes. The results are satisfactory in light of the available data, from the aggregate initial demand matrix to the link counts used, and provide encouraging indications for the ability of the DTA tool to support the intended analysis of traffic patterns under various scenarios. Figure 4.12 depicts an example of the results of the simulated link volumes compared with observed link volumes for a selected link.

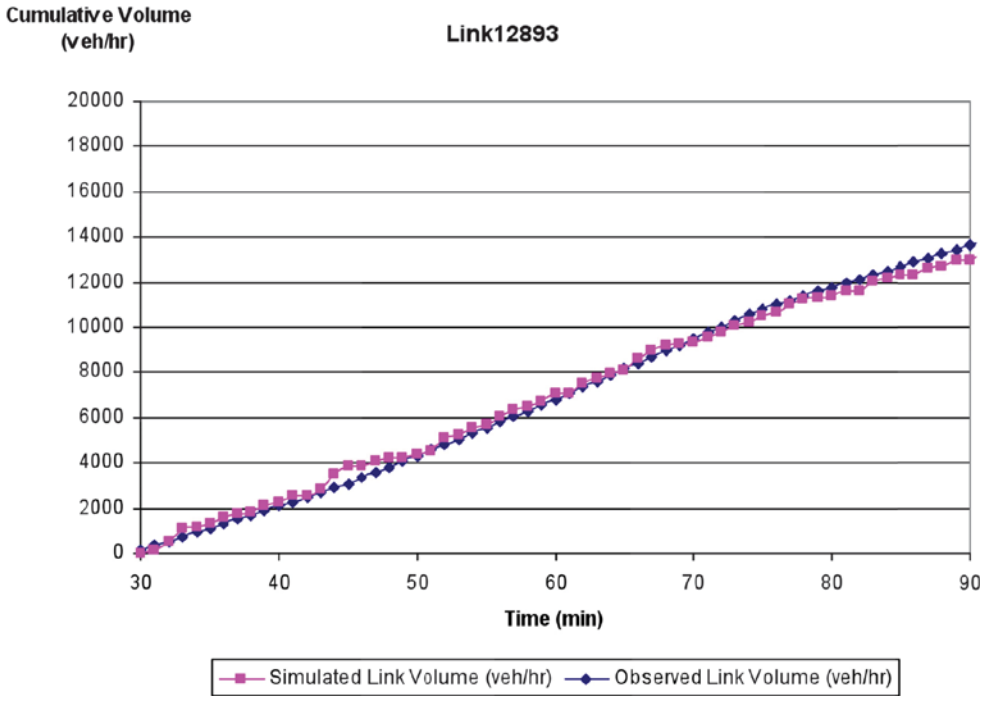

Figure 4.12. Sample of simulated link volumes versus observed link volumes.

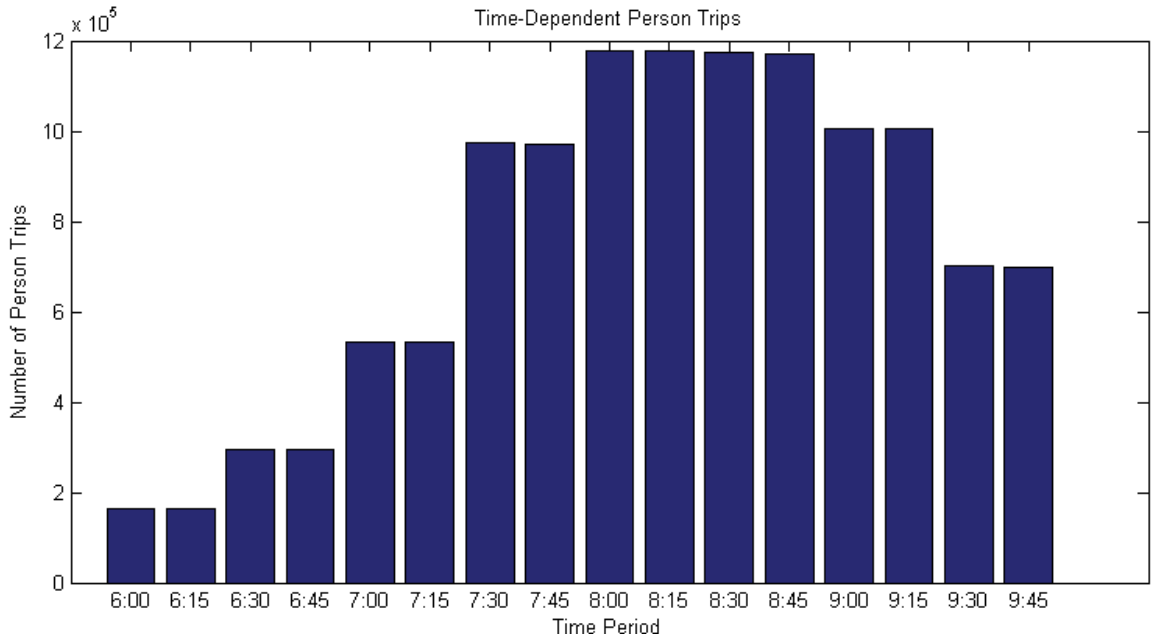

Figure 4.13. Time-dependent person trip departures.

Numerical Experiments

Scenario Definition

The planning horizon is the morning period from 6:00 to 10:00 a.m. The departure time interval is 15 minutes. Figure 4.13 shows the time-dependent pattern of person trip departures in this study.

Link toll information is obtained based on the existing pricing schemes implemented in the New York region. Six of the 297 tolled links are dynamic toll roads for peak periods (weekdays, 6:00–9:00 a.m. and 4:00–7:00 p.m.; Saturday and Sunday, noon–8:00 p.m.). The price distribution applied for tolled links is given in Figure 4.14, and the road type distribution is shown in Figure 4.15.

The experimental set-ups for the MDUE model were established with the aim of validating the performance of the gap-based selection technique in the bicriterion time-dependent shortest-path calculation and evaluating the solutions with different settings of VOT step sizes (Δ). Three experiments were set up as presented in Table 4.1.

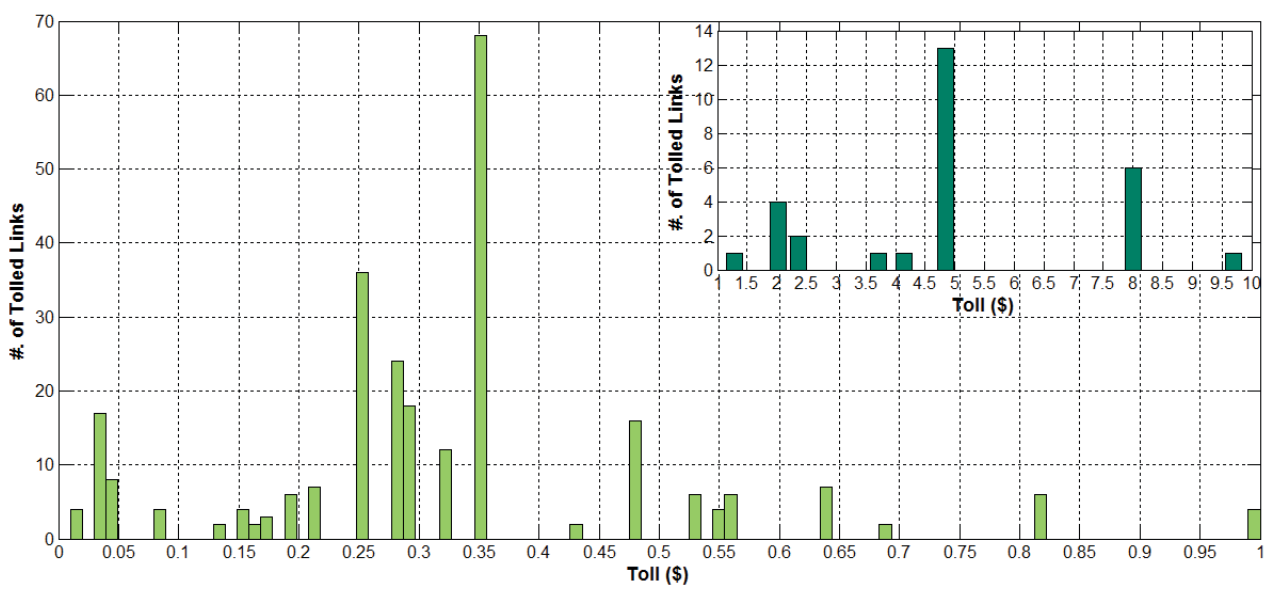

Figure 4.14. Price distributions for tolled links in the New York network.

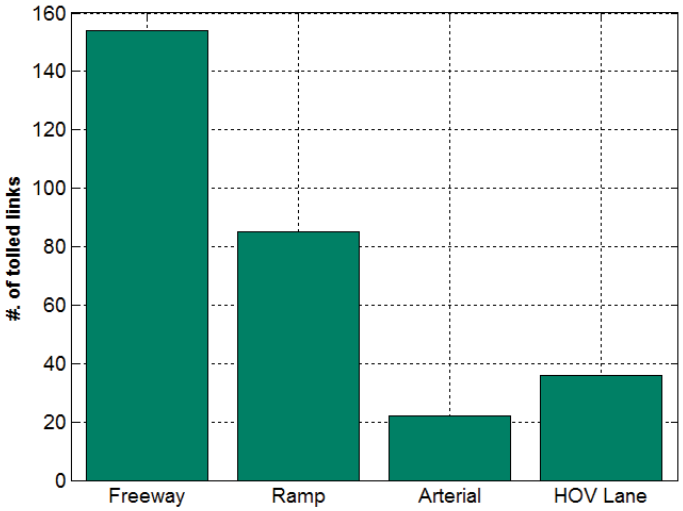

Figure 4.15. Road-type distributions for tolled links in the New York network.

Convergence Pattern of the Integrated Model

The convergence of the algorithm is examined by the objective function of formulation described in Appendix A, specifically in Equation A.16, for the dynamic mode choice problem. This expression is a gap measure of the total square of the difference between assigned mode flows y_m^{wt} and expected mode flows $q^{wt} \times p_m^{wt}(y)$ calculated by the mode choice model under the prevailing trip times and network attributes obtained with those assigned flows y_m^{wt}. This measure is an extension of the convergence criterion defined in Zhang et al. (2008) for a generalized dynamic SUE problem. Figure 4.16 shows the convergence pattern in terms of this average gap measure.

Table 4.1. Experimental Set-up for MDUE Model

	Gap-Based Technique	VOT Step Size (Δ)	Outer Loop	Inner Loop
Experiment 1	No	0.3	5 Iterations	1 Iteration
Experiment 2	Yes	0.3	5 Iterations	1 Iteration
Experiment 3	Yes	0.5	5 Iterations	1 Iteration

The smaller this quantity is, the closer the agreement will be between the demand values corresponding to the travel time attributes (by the mode choice equation) and the assigned flows that have produced these network attributes. In other words, smaller values reflect consistency between the demand models and network performance simulation, thereby corresponding to an equilibrium solution that satisfies the TDMSUE conditions defined in the previous section. The convergence pattern exhibited in Figure 4.16 suggests successful equilibration and a very efficient overall iterative scheme.

Convergence Pattern of Multicriterion Dynamic Use Equilibrium Model

From a methodological perspective, the convergence patterns of the proposed MDUE solution algorithm together with different implementation techniques are examined on the New York metropolitan regional network. The objective function $Gap(r) = g_R(x, \alpha_I)$ [where $g_R(x, \alpha_I)$ is as defined in Equation A.41 in Appendix A] is calculated based on vehicle-experienced generalized cost at each iteration. Another measure

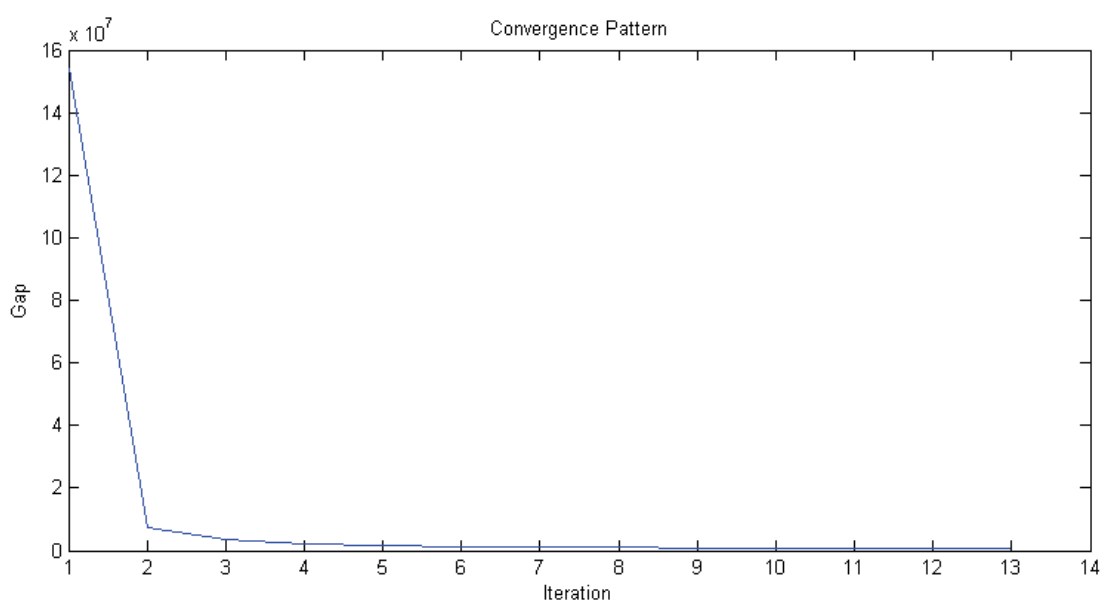

Figure 4.16. Convergence pattern of TDMSUE solution.

of effectiveness is collected in all conducted experiments, in addition to the objective function $Gap(r)$. The additional measure is the average gap over all the vehicles in the network for a given path flow pattern r:

$$AGap(r)=\frac{\sum_w\sum_t\sum_m\sum_k\sum_i x_k^{wtm}(\alpha_i)\times[GC_k^{wtm}(x,\alpha_i)-\pi^{wtm}(x,\alpha_i)]}{\sum_w\sum_t\sum_m\sum_i \tilde{K}^{wtm}(\alpha_i)} \tag{4.10}$$

$AGap(r)$ is used as a surrogate of the gap function $Gap(r)$ and is calculated based on the vehicle-experienced path generalized cost in this study. The lower bound of the $AGap(r)$ is zero. Essentially, the smaller the average gap, the closer the solution is to an MDUE, as differences in generalized cost across paths used at equilibrium between a given O-D pair at a given time become very small. To get a better illustration of the solution quality, $AGap(r)$ is calculated with and without considering travel time reliability in the path generalized cost function and is reported separately, as shown in Figure 4.17.

As shown in Figure 4.17 (top), the MDUE algorithm can effectively reduce the average gap measures (with travel time reliability) in all three experiments, although the convergence patterns are not strictly monotonically decreasing. As for solution quality, the final average gap values at least iteration reduced 63.56%, 64.93%, and 62.25% of the initial gap values, respectively, for the three experiments. Additional convergence criteria based on average gap without travel time reliability, as shown in Figure 4.17 (bottom), exhibit similar convergence trends, with relatively lower average gap values obtained at each iteration as an additional term is left out in the path generalized cost. Recognizing the complexity of the problem and the scale of the network, the convergence patterns indicate that the MDUE algorithm can find a sufficiently close-to-MDUE solution for the New York metropolitan regional network.

Computational Time Analysis of Multicriterion Dynamic Use Equilibrium Model

From a practical application standpoint, the effectiveness of the implementation techniques is investigated in this section. As discussed above, the most intensive computational operation in MDUE solution algorithm is PAM, which calculates the time-dependent least-generalized cost TDLGC path tree and partitions the entire feasible range for each destination. The computational time required by MNDL in DYNASMART and RMDUE (inner loop) only depends on the total number of vehicles loaded on the network; therefore it remains constant at any iteration in all experiments. Actual times depend on the specific hardware configuration used, although the former is in the order of 2 hours and the latter 2 minutes on medium-end workstations.

To examine the effectiveness of the implementation techniques, the relative computational times required by PAM in three experiments are demonstrated in Figure 4.18. Relative

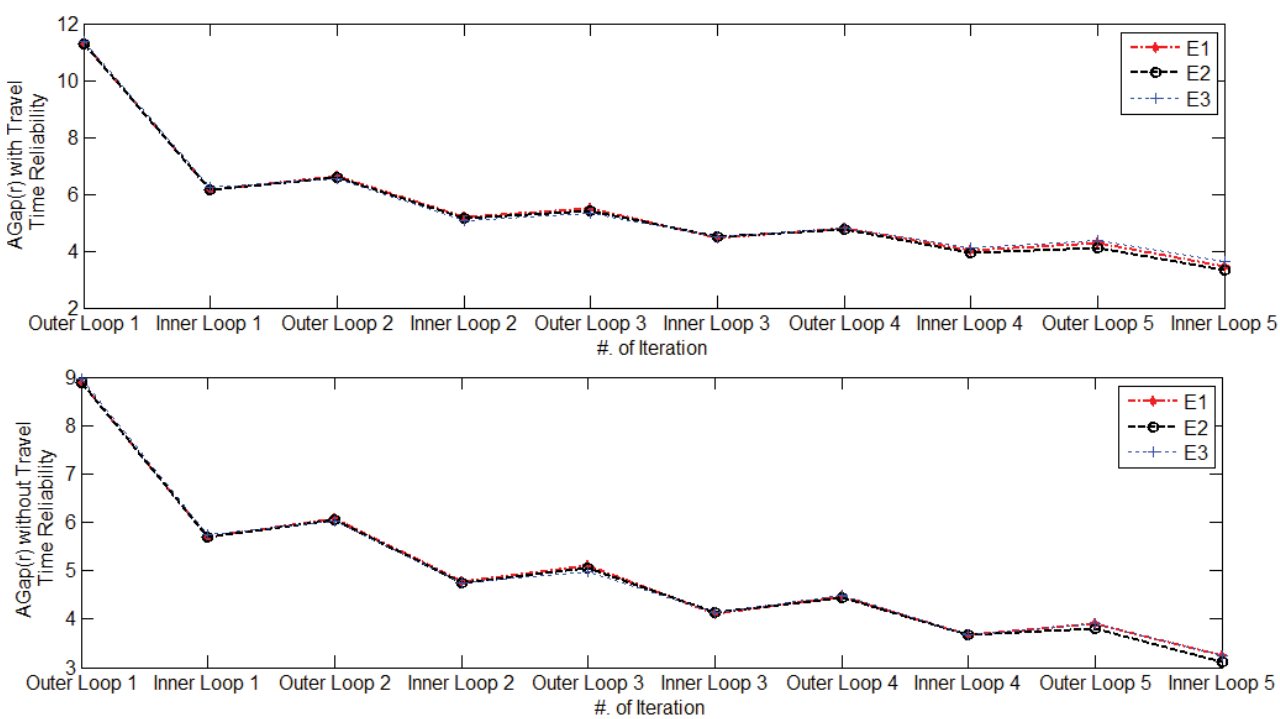

Figure 4.17. Convergence patterns in terms of average gap* (top) *with travel time reliability and* (bottom) *without travel time reliability.

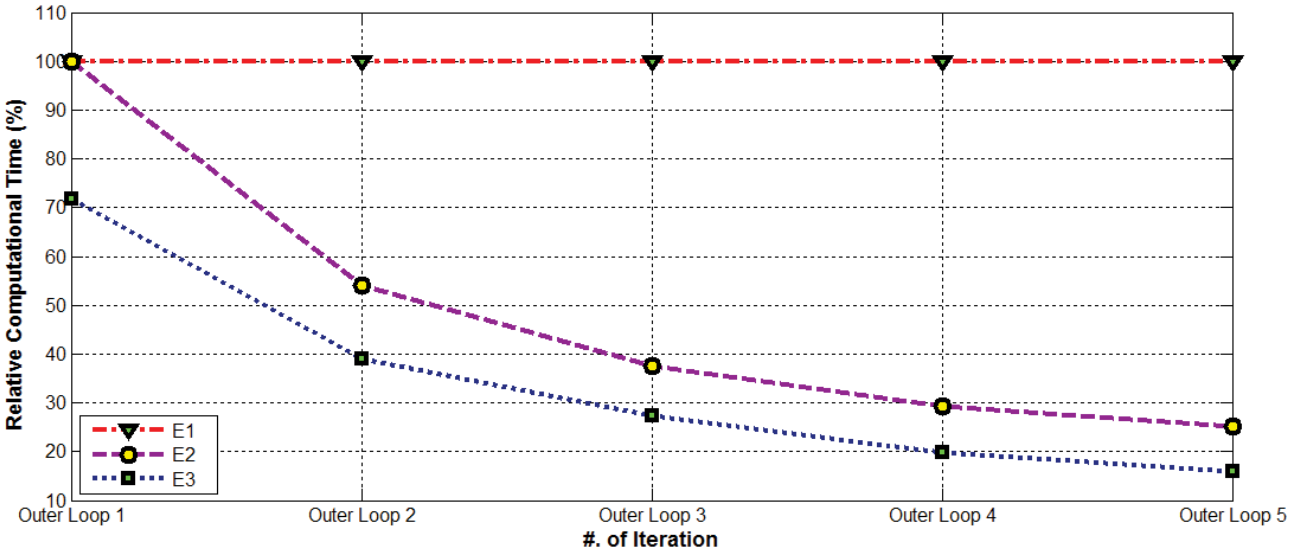

Figure 4.18. Computational time required by PAM (parametric bicriterion path-finding procedure).

computational time is defined as the average time required by one operation in PAM for a root node (destination) of a TDLGC path tree relative to the average time in E1; this measure is used to explore the magnitude of the improvement. The amount of time required in PAM depends on the number of destinations and VOT step size. The full run of PAM with 0.3 for VOT step size in E1 requires a large number of computations and takes the most computational time. In E2, a gap-based selection technique was implemented that reduced the number of TDLGC path trees in PAM by a factor of $1/k$ at each iteration. There was an approximately 50% reduction of total time in E2 compared with E1. The improvement was the most significant in E3, in which both a gap-based selection technique and adjustment of VOT step size to a greater value (0.5) were implemented, because the number of computations involved was lower relative to other experiments. It was also observed that the marginal contribution (by gap-based selection technique) of reduction on computational time diminishes as the number of iterations increases, as those destinations left in PAM are the most sensitive to VOT value and thus require time to partition VOT ranges. Both E2 and E3 attained comparable levels of solution quality (see Figure 4.17), but PAM in E3 was faster (i.e., only 72% of the total time required in E2). Therefore, the combined implementation techniques provide the most reduction in computational time for the MDUE solution algorithm.

Analysis of Mode Choice Results

According to the NYBPM model and the latest calibrated nested logit mode choice model, there are 13 modes in this study, as depicted in Figure 4.3. This study specified three income categories: low (average income is $7,182), middle (average income is $41,065), and high (average income is $129,795); this stratification is based on the 1990 PUMS data. In addition, there are eight trip purposes including work with low income, work with middle income, work with high income, school, university, maintenance, discretionary, and at-work.

The mode share pattern was analyzed by time period, trip purpose, and income level. Figure 4.19 shows a time-dependent mode share pattern of travelers in which travelers are inclined to drive alone in the early morning. That is, 6:00–7:30 a.m. is the peak period for SOV travelers; travelers will choose ride-sharing after 7:30 a.m. The share of toll road users does not change very much in this case study, partly because the number of toll links is very small in the network. The shares of transit and rail modes increase slightly from 6:00–10:00 a.m.

Figure 4.20 shows mode shares by trip purpose. It can be seen that more than 40% of high-income travelers will drive alone to work, which is also confirmed by Figure 4.21, which shows mode share by income group. Similarly, low-income travelers tend to use transit modes to go to work.

Figure 4.22 shows the time-dependent vehicle patterns by occupancy level (drive alone, HOV2, HOV3, and HOV4+). From this figure, it can be seen that the time-dependent pattern is consistent with the time-dependent mode share pattern in Figure 4.19, which confirms the correct integration of dynamic mode choice and route choice model in this study.

Impact of Implementation Techniques and Continuously Distributed Value of Time of Multicriterion Dynamic Use Equilibrium Model

An important additional consideration is whether using the implementation techniques would lead to large differences in

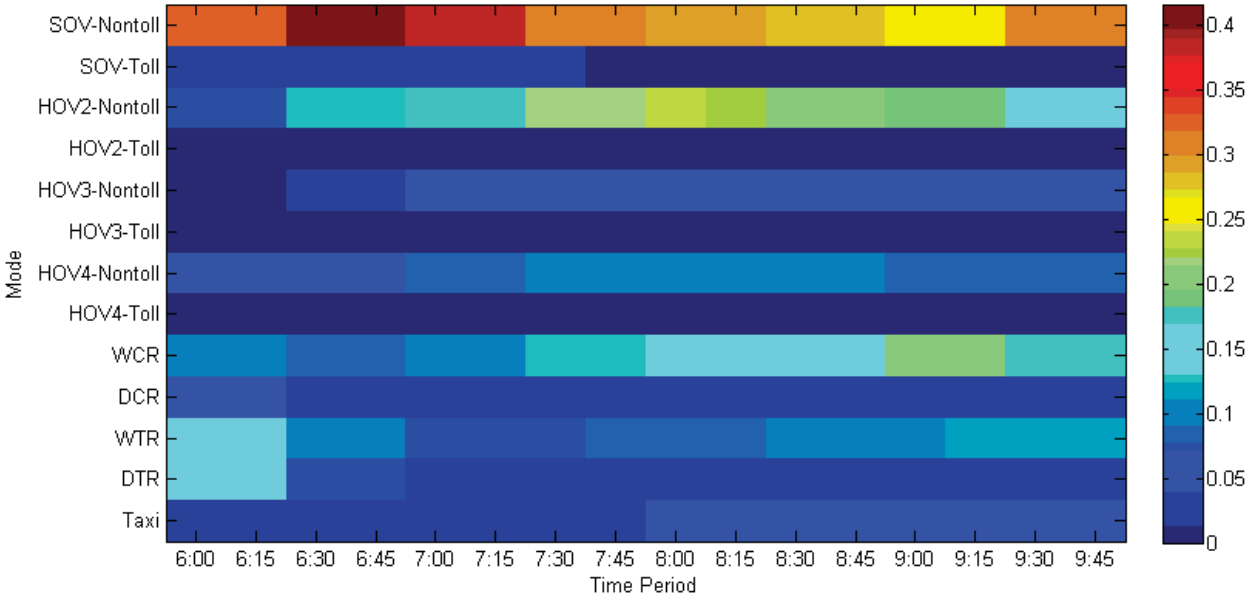

Figure 4.19. Time-dependent mode share.

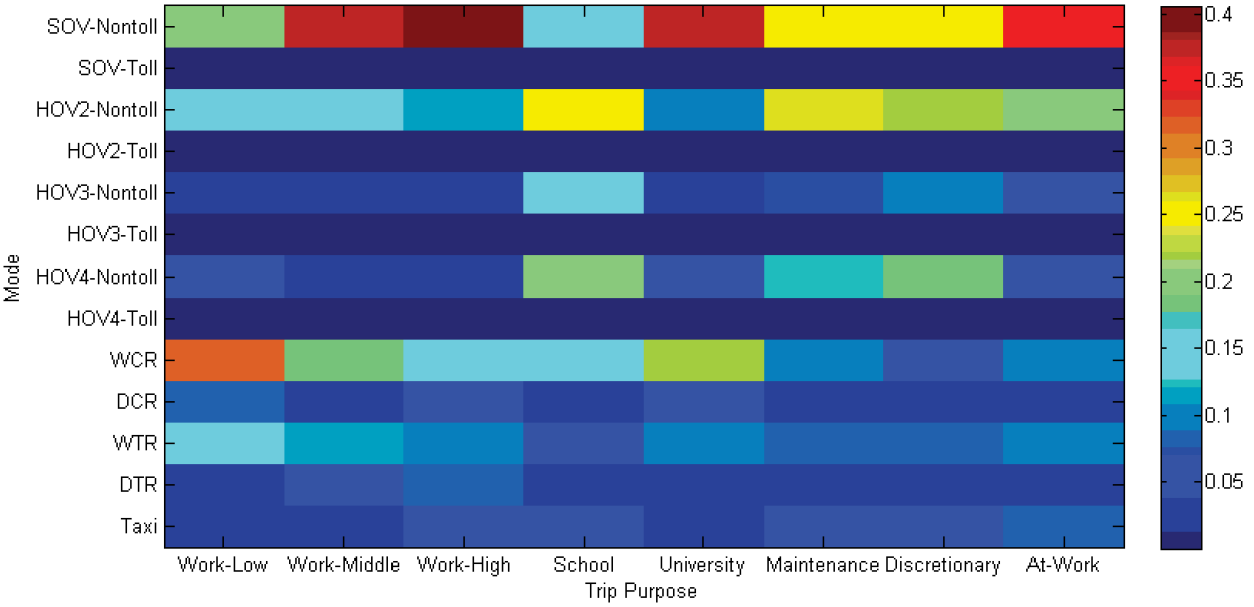

Figure 4.20. Mode share by trip purpose.

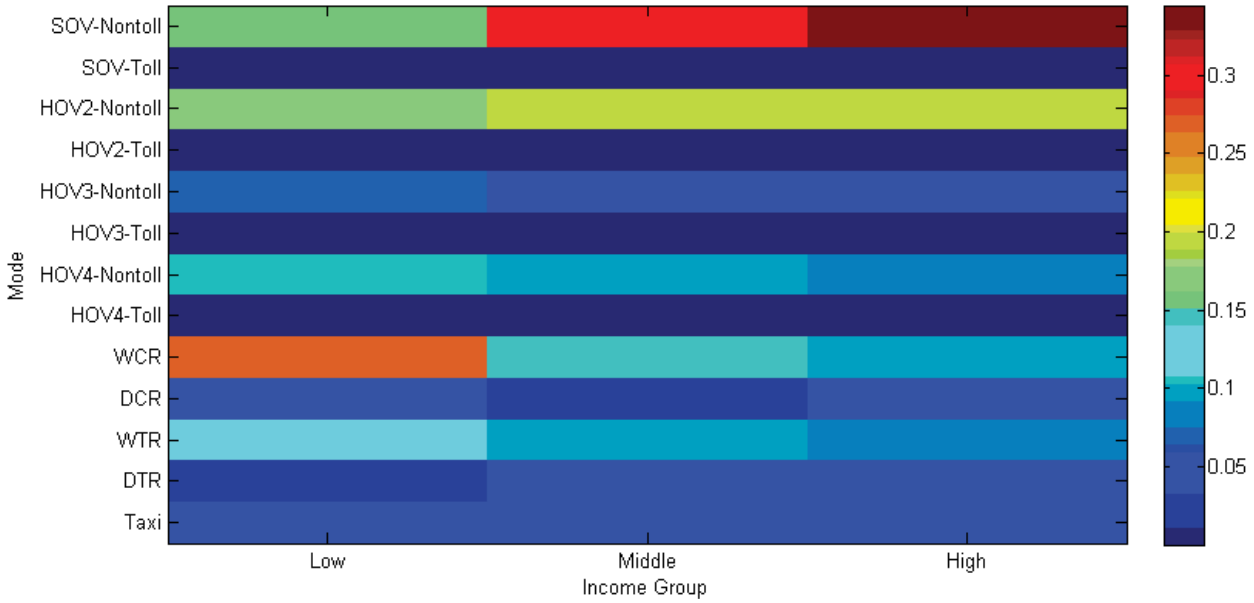

Figure 4.21. Mode share by income group.

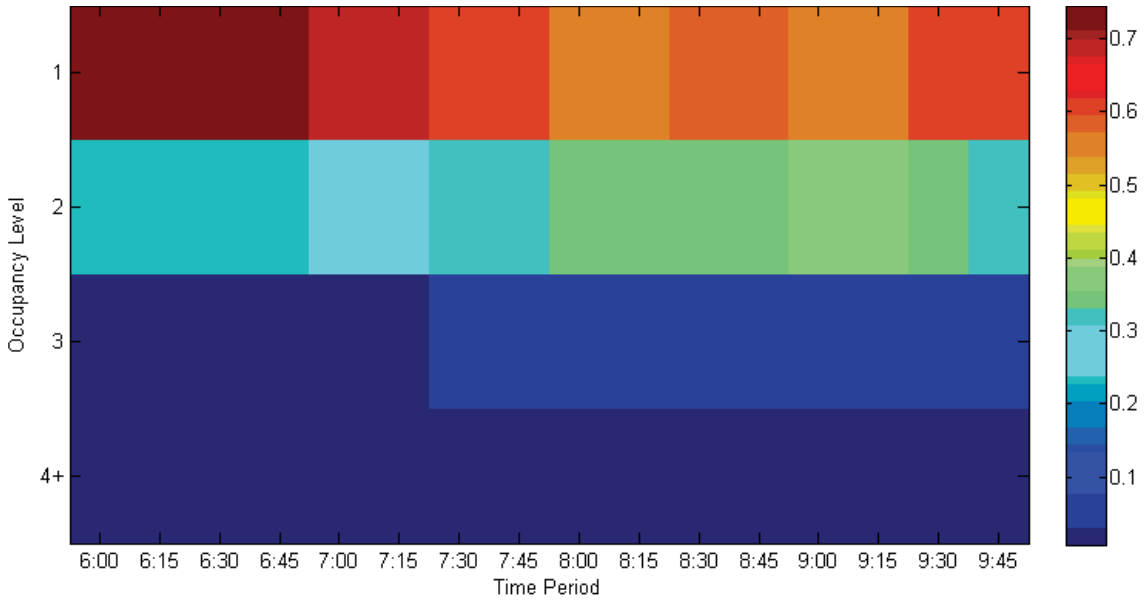

Figure 4.22. Time-dependent vehicle share by occupancy level.

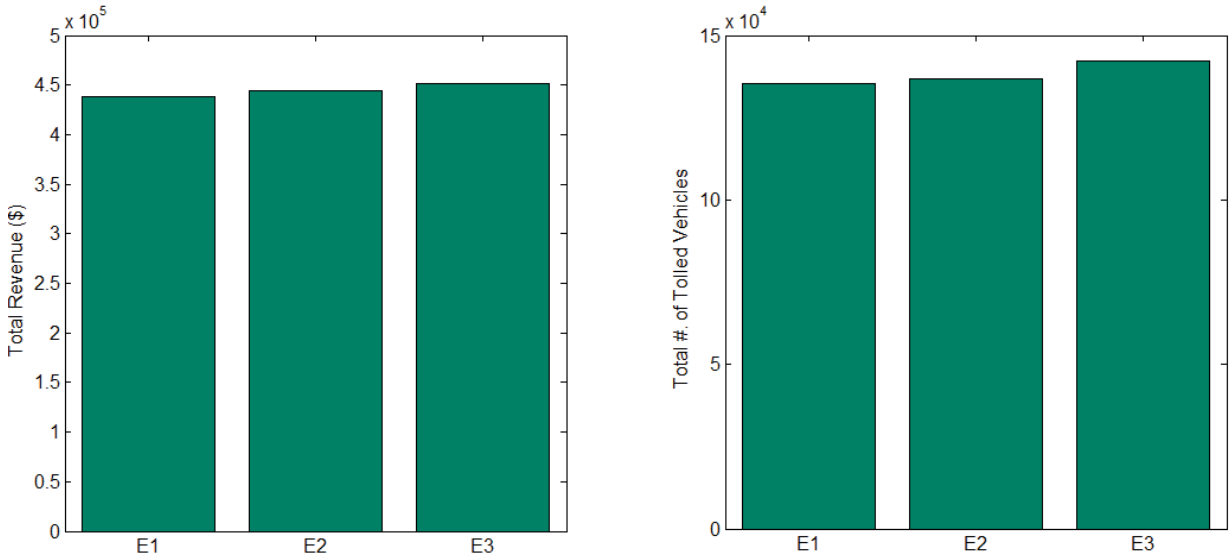

Figure 4.23. Toll road usage for* (left) *total revenue and* (right) *total number of tolled vehicles.

prediction of the flow patterns on the network. To address this concern and investigate the impact of continuously distributed VOT across the entire population, numerical results regarding the toll road usage are presented in this section. Toll road usage is examined from three perspectives: total revenue, total number of vehicles passing through toll links, and to reflect heterogeneity, grouped toll road users in different VOT segments.

It is observed in Figure 4.23 that the stability of toll road usage (both in terms of total revenue collected and total number of vehicles passing through the toll links) is attained in all experiments and stays at approximately the same level. Figure 4.24 provides the toll road usage of trip makers in different VOT segments over the planning horizon predicted in all experiments. Note that the partition of the feasible range of VOT distribution ($[\alpha^{min}, \alpha^{max}]$) is independent of the step size selected in the MDUE algorithm. As illustrated in Figure 4.24, trip makers with different VOT react differently to a given road pricing scheme; thus significant discrepancies are obtained under the conventional assumption of homogeneous (constant VOT) users. Toll road usage is lower for those trip makers in the low VOT segment, but higher for high VOT users. Again, all experiments predict similar proportions of toll road users in all VOT segments. Supported by those results, it can be concluded that the MDUE algorithm can capture greater realism in path choice behavior, and the proposed

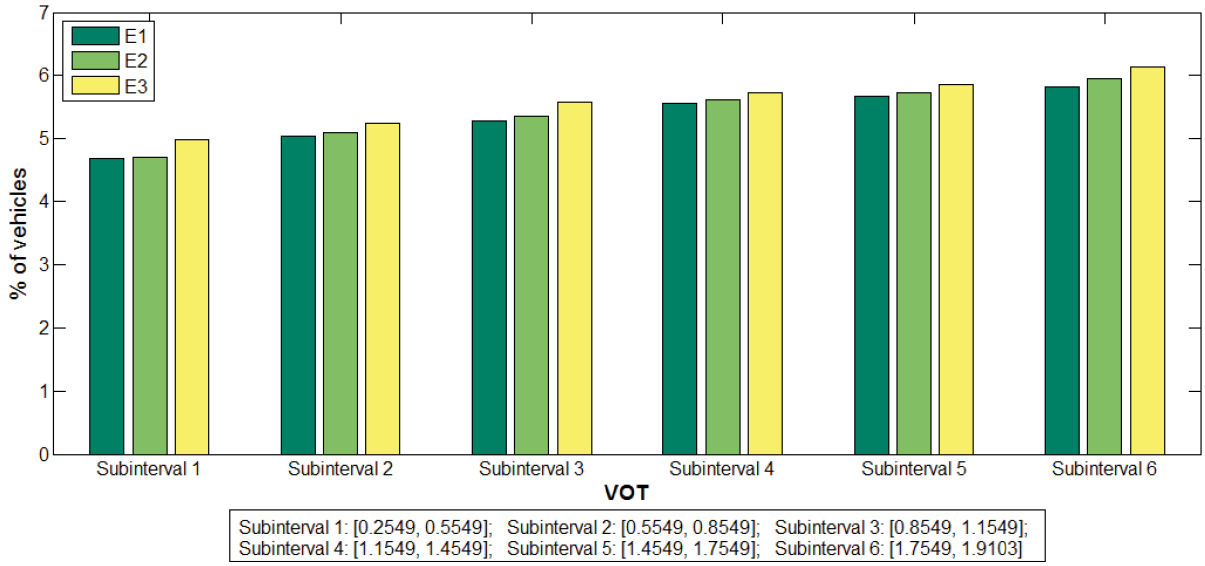

Figure 4.24. Grouped toll road users in different VOT segments.

implementation techniques do not compromise the accuracy of flow pattern prediction.

Summary of Network Modeling Procedures

The proposed integrated model framework is a demonstration of a trip-based integration of a well-calibrated mode choice model in practice and a simulation-based dynamic traffic microassignment model. However, the framework is sufficiently flexible to incorporate other dimensions (e.g., destination choice and departure time choice) in addition to the mode choice dimension from the demand side. In addition, the framework can be readily extended to an activity-based integration of demand models and an activity-based dynamic traffic microassignment model. The basic methodological challenges inherent in such integration problems have been substantially addressed and demonstrated in the present study. Extension and generalization would require additional investment in time and effort, as well as considerable detail in implementation that would be specific to the activity demand structure developed in a particular area, albeit using and building on components demonstrated in the present study. Accordingly, this study provides an essential foundation and direction toward an evolution that provides a conceptually, theoretically, and methodologically complete and sound approach to address heterogeneous user responses to congestion, pricing, and reliability in large-scale regional multimodal transportation networks.

The principal contributions of the integration effort presented in this chapter include the following:

1. Integrating additional dimensions of user choice in a network simulation assignment platform;
2. Consistently solving a complex realistic stochastic mode choice model equilibration problem in connection with a large-scale network assignment process;
3. Incorporating multiple attributes in the route choice process, particularly price and travel time reliability, in addition to mean travel time;
4. Recognizing user heterogeneity in terms of preferences, especially with regard to VOT in response to pricing schemes and other congestion-related measures;
5. Calibrating a route choice model with a distributed VOT following a lognormal distribution;
6. Devising and testing path-finding procedures that recognize multiple criteria and developing efficient implementations for large-scale networks;
7. Incorporating travel time reliability in the route choice process and devising a robust and efficient traffic-theoretic procedure to generate reliability attributes for path-level choices; and
8. Demonstrating the integrated procedure on the actual network of the New York metropolitan region, which is the largest application of DTA equilibration procedures reported to date.

In the application to the New York region, the team presented dynamic mode share and toll road usage results of the proposed integrated model for the region's networks. These results demonstrated the applicability of the model and procedures developed in this work to practical large-scale networks. The team also examined the convergence of the proposed algorithms, establishing successful attainment of the desired equilibrium conditions at all levels of the procedure in connection with both route and mode choices. The convergence process revealed a relatively efficient iterative process, further supporting the practical applicability of the integrated procedures developed in this work. The proposed model and the implementation techniques uniquely address the needs of metropolitan areas and agencies for prediction of mode and path choices and the resulting network flow patterns and provide the capability for evaluating a wide range of road-pricing scenarios on large-scale networks.

CHAPTER 5

Incorporation of Results in Operational Models in Practice

This chapter provides a concise guide to how the methodological issues and results presented in the previous chapters can best be incorporated into practical planning tools. Specific models and tables presented in earlier chapters are referred to as needed.

Trip-Based Four-Step Demand Framework

In general, the four-step aggregate zonal framework is not sufficiently flexible to fully incorporate most of the model specifications tested in this project. However, a number of advances can readily be incorporated:

- Segment the zonal population by income group. The effect of pricing on behavior is strongly related to income. In order to capture the effects of pricing somewhat realistically, the population of each zone should be segmented into at least three or four income groups, and income should be included as a variable in all models, including mode choice, destination choice, trip generation, and auto ownership. The mode choice model results in described in Chapter 3 recommend specifications for including income as a modifier for the cost variable. Separate mode-specific constants can also be estimated for different income groups.
- Use different auto paths for different income groups. If there are priced links in the auto network, then the shortest path through the network in terms of generalized time or generalized cost will vary with value of time (VOT), which in turn varies by income. For each income group, a VOT similar to the one used in the mode choice models should be used to calculate the generalized time along each path, and separate shortest-path skim matrices of travel time and toll cost should be created for each income group. This approach is used in the Puget Sound Regional Council (PSRC) trip-based models.
- Include a toll–nontoll choice submodel in the mode choice. A toll–nontoll choice can provide more flexibility and realism to skim the best tolled path and the best nontolled path separately and model that choice as a nested, binary subchoice in mode choice models. The New York mode choice models presented in Chapter 4 use such a structure, which can be used for either trip-based or tour-based models. In this case, different tolled path skim matrices should be prepared for each income group, but the best nontolled path can be the shortest time path, which is the same across income groups.
- Include the effects of pricing and congestion and alternative modes in trip distribution and destination choice models. As described in Chapter 4, there are more comprehensive measures to use in location choice models than simply using the travel time by auto, which is often done in gravity-type distribution models. A better form is the inclusive value, or logsum, across all travel modes, which includes the effects of both pricing and congestion, and also takes into account accessibility by nonauto modes. The travel times and costs used in calculating the logsums should be for the most representative periods of the day for the trip purpose (unless, in the ideal case, the logsum is from a joint trip mode and time-of-day [TOD] choice model, as presented in Chapter 3, in which case all time periods of the day will be included in a representative way). Ideally, the logsum will be included as an impedance value in a discrete destination choice model. Several existing four-step model systems use destination choice models in place of gravity models, at least for the commute purpose (e.g., the Southern California Association of Governments and PSRC models). Even with a gravity-type model, however, it is still possible to normalize a mode choice logsum to use as an impedance variable, such as by dividing by the travel time coefficient to convert it into equivalent minutes of travel time.
- Include an explicit auto ownership model and segment subsequent models by auto availability. Although auto

ownership models were not estimated in this study, they are included in most advanced-practice trip-based models, and examples can be found in the literature. The team includes this recommendation here to underline its importance, because the most significant variables in mode choice models are those related to auto ownership, and auto ownership is responsive to changes in household size and income distributions, as well as changes in pricing and congestion. Trip generation, distribution, and mode choice models should then be segmented by both income and auto ownership, with auto ownership divided into at least three segments: (1) households with zero autos, (2) households with one or more autos but fewer autos than working adults, and (3) households with one or more autos per working adult. The mode choice models presented in Chapter 3 show the importance of those segmentation effects. Note that for nonwork trips, it may be best to define the segments in terms of autos per driving-age adult rather than per working adult.

- Include accessibility variables in the auto ownership and trip frequency models. One major criticism of four-step models is that major changes in travel congestion or prices do not affect the upper-level models, such as trip generation, so they are not able to predict induced or suppressed trips. Chapter 3 describes how to specify accessibility variables to use in upper-level choice models. Although they are discussed in the context of activity-based models (ABMs), there is no reason why they cannot be used in four-step model frameworks, particularly as they are aggregate, traffic analysis zone–based measures defined for a limited number of specific population segments. Note that in order to include such variables in a trip frequency model, a regression model (Poisson regression is appropriate for count data) will be needed, rather than a simple cross-classification table.
- Include explicit TOD choice models, ideally as joint models with mode choice. Most four-step models in practice use fixed TOD factors that are not sensitive to the relative travel speeds and prices in different periods. To predict peak-spreading phenomena, explicit TOD choice models are needed. Although TOD models work best as activity and trip scheduling models within a tour-based, full-day model framework, they can still provide substantial benefit at the trip level, particularly to model the effects of TOD pricing. Using the basic specifications in Chapter 3 for when the TOD model is applied to car trips according to mode choice or when the model is estimated and applied jointly with a mode choice model, the TOD model provides the basic sensitivity TOD variations in pricing and congestion. Note that a TOD model will work best with at least five network assignment or skim periods for auto: pre-a.m. peak, a.m. peak, midday, p.m. peak, and post-p.m. peak. A TOD model can benefit from the use of even more periods, such as separate skims for the shoulders of the peak. Conversely, it would not be beneficial to include skim matrices for more times of day without also including a TOD choice model.

As more and more of the improvements recommended above are incorporated into the aggregate trip-based framework, the inefficiencies in terms of computation and run time can become extreme. In particular, the amount of computation in the zonal aggregate framework increases linearly with the number of population segments, and it can also increase substantially with the number of TOD periods (particularly if separate network skims are required for different income groups). Also, the full computation of mode choice logsums across all possible destination zones becomes prohibitive with large numbers of zones. In contrast, in microsimulation model frameworks in which each household is simulated, the run time does not increase substantially with the number of household and person variables in the models, nor with the number of time periods or the number of zones (particularly since sampling of destinations can be used). As a result, the more advanced features one wishes to incorporate into a trip-based four-step model, the more beneficial it becomes to move to an activity-based microsimulation framework, in which even more advanced features can readily be incorporated, as described below.

Advanced Tour-Based, Activity-Based Demand Models

Many of the recommendations for tour-based ABMs parallel the ones provided for the trip-based four-step models described above. For activity-based microsimulation models, however, the possibilities are greater, and the issues tend to be somewhat different. As described in Chapter 6, some of the issues may depend on whether the demand models are to be applied in combination with static or dynamic network assignment methods. For the SHRP 2 C10 project, for example, ABMs will be applied in combination with the Dynus-T and TRANSIMS network models. The same ABMs can (and will) also be applied in combination with more conventional static equilibrium assignment methods; however, the desired capabilities may differ.

For advanced tour-based ABMs, the team recommends the following model specifications:

- Represent as many determinants of VOT as possible, including residual heterogeneity (simulate a specific VOT for each person and tour). With ABMs applied with an agent-based microsimulation framework, it is possible to include a large variety of household, person, and trip characteristics in the models. The analyses in Chapter 4 indicate systematic

effects of a number of those characteristics on willingness to pay and VOT. A prototypical VOT function would be one such as in Chapter 3, including the effects of income, occupancy, mode, TOD, gender, age, tour purpose, and distance, plus residual random variation. Since the latter term is represented as a standard deviation of a lognormal distribution, then applying such a model in practice involves the following steps: (1) for a given tour and trip made by a given person, apply the VOT coefficients for the relevant household, person, trip, and tour characteristics to calculate the systematic portion of the cost and time coefficients; (2) use a random draw to stochastically draw the random portion of the time coefficient given the distribution and the standard deviation; and (3) combine the systematic and stochastic components of the time and cost coefficients to calculate a person- and tour-specific VOT to use in the choice models for that simulated individual's travel.

- Use different auto paths for different VOT groups. If there are priced links in the auto network, then the shortest path through the network in terms of generalized time or generalized cost will vary with VOT. In contrast to the recommendation for trip-based models, for which income segmentation is the main determinant of VOT, an ABM that simulates a person- and tour-specific VOT (as recommended above) will have a specific VOT associated with every trip. So, the best strategy for producing different skims to feed back to the models is to segment the trips within different VOT groups (e.g., by quartiles). For each VOT group, the median VOT within that group should be used to calculate the generalized time along each path, and separate shortest-path skim matrices of travel time and toll cost should be created for each income group.
- Include a toll–nontoll choice submodel in mode choice. As recommended for trip-based models, including a toll–nontoll choice can provide more flexibility and realism to skim the best tolled path and the best nontolled path separately and to model that result as a nested binary subchoice in mode choice models at the trip level, and possibly at the tour level. The New York mode choice models presented in Chapter 4 use such a structure, which can be used for either trip- or tour-level models. Different tolled path skim matrices should be prepared for each VOT group, as described in the preceding paragraph, but the best nontolled path can be the shortest-time path, which is the same across VOT groups.
- Include the effects of pricing and congestion and alternative modes in tour and trip destination choice models. As described in Chapter 3, the inclusive value, or logsum, across all travel modes, which includes the effects of both pricing and congestion and also considers accessibility by nonauto modes, should be used in any tour or intermediate-stop destination choice models. The travel times and costs used in calculating the logsums should be for the most representative periods of the day for the tour or trip purpose (unless, in the ideal case, the logsum is from a joint trip mode and TOD choice model, as presented in Chapter 4, in which case all time periods of the day will be included in a representative way).
- Include accessibility variables in the upper-level auto ownership and tour frequency and activity pattern models. Chapter 3 describes how to specify accessibility variables to use in upper-level choice models, which have been used in a number of recent ABM systems. This strategy should be followed as closely and comprehensively as possible in order to include consistent effects of pricing and congestion at all levels of the model system. Most ABM systems are sufficiently flexible to include such accessibility variables in a variety of ways. Even if the effects of variables are only marginally significant in model estimation, it is still advisable to include them as long as the signs are correct. Note that there is not yet enough evidence for these variables to recommend specific coefficient values to use in general cases, although that may be possible in the near future after a number of additional ABM systems have been estimated.
- Include explicit TOD choice models, ideally as joint models with mode choice. TOD models work best as activity and trip scheduling models within a tour-based, full-day model framework. The hybrid departure time and duration specification described in Chapter 3 is recommended, with TOD and duration shift effects. Such models can be at both the tour level and the trip level, using the specifications of Chapter 3, whether the TOD model is applied to car trips according to mode choice or the model is estimated and applied jointly with a mode choice model. Note that with static assignment procedures, a TOD model will work best with at least five network assignment or skim periods for auto: pre-a.m. peak, a.m. peak, midday, p.m. peak, and post-p.m. peak. A TOD model can benefit from even more periods, such as separate skims for the shoulders of the peak.

With dynamic traffic assignment (DTA), the level of service (LOS) can be fed back for a large number of time periods, since it is typical for those methods to work at a fine level of temporal detail. The methods for passing back LOS from a dynamic traffic simulation are discussed in the following section.

Static and Dynamic Traffic Simulation Tools

As explained in the beginning of this chapter, static network modeling tools used in conjunction with four-step demand forecasting and planning tools are not especially well suited to capture user responses to dynamic pricing schemes, congestion, and (un)reliability. Recommendations for improvements

in the application of these methods (see beginning of this chapter) include the main recommendation of segmentation of the demand matrices into classes with different VOT and other behavioral parameters. Segmentation is a commonly used approach to incorporate more demand-side realism in four-step aggregate procedures. It does not, however, accomplish much in terms of a more realistic representation of the supply side, particularly congestion dynamics and reliability. In addition, the user class definitions are subject to a certain degree of arbitrariness, in addition to containing users with potentially distinct network path and mode choices. Furthermore, it adds considerably to the computational burden of applying these procedures.

Most of the interest therefore lies in dynamic, simulation-based tools for network assignment. Microassignment methods, which track individual particles, have become the state of the art in DTA models advancing to the early stages of practice. These methods provide a natural platform for incorporating the kinds of behaviorally rich demand-side tools developed as part of this project.

The challenges in integrating individual-level ABMs and the pricing- and congestion-responsive behavior models developed here are discussed in detail in Chapter 4. Solutions to these challenges are developed and demonstrated in that chapter, in connection with a particle simulation-based dynamic equilibrium assignment methodology. As demonstrated, these methodological innovations are ready for implementation in connection with existing simulation-based network assignment tools. Note in this regard that the differences in the physics underlying the traffic propagation (simulation) per se (e.g., the traditional distinctions between micro- and mesosimulation approaches) are not directly relevant to the applicability of the network procedures developed and demonstrated as part of this work. The main requirement is that the approach represents and tracks individual travelers as decision entities. Beyond that, whether vehicle propagation invokes robust and easy-to-calibrate relations among averages to determine speeds at which vehicles move, or detailed microscopic rules for car-following and passing maneuvers, is not essential to the procedures developed under this project.

Four main modeling elements introduced in this work are essential to accomplish the objectives of the study, namely, to develop predictions of network flows and facility usage while capturing user responses to various forms of pricing, congestion, and reliability. These four elements, which are outlined in Figure 5.1, form the recommendations for advancing the state of the practice by implementing the methods developed for this project:

1. Expand the set of attributes typically considered in the route choice assumptions underlying traffic assignment methods beyond travel time to include cost (toll) and

CHALLENGES	HOW ADDRESSED	EFFORT REQUIRED FOR USE IN CONNECTION WITH EXISTING METHODS
EXPAND SET OF LOS ATTRIBUTES IN ROUTE CHOICE AND OTHER CHOICE DECISIONS	Network attributes expanded in generalized cost (disutility) path search and assignment	Relatively simple to implement in path-based assignment methods
CAPTURE HETEROGENEITY OF TRAVELERS	Two approaches: (1) User Class Segmentation vs. (2) Continuously Distributed Value of Time with parametric shortest path	Given estimated distribution, implementation of approach 1 possible but time-consuming to execute. Approach 2 much more general but requires greater code intervention.
GENERATE RELIABILITY MEASURES AS LOS ATTRIBUTE	SHRP-2 L04 developing comprehensive methodology; Immediately applicable; relation between std. dev. and mean travel time, both per unit distance	Immediately applicable relation easy to establish and implement at path level in many DTA tools.
COPE WITH LARGE-SCALE NETWORK EXECUTIONS ISSUES	Novel algorithms and implementations	Requires code-level intervention and improvement; difficulty and level of effort would depend on existing implementation and data structures.

Figure 5.1. Summary of major network issues and proposed solutions.

reliability measures in the form of generalized cost. Most existing tools readily allow incorporating a link-level cost attribute in the path search procedures used to generate paths in the assignment process. However, incorporating reliability is more challenging, because most measures of reliability are not additive across links. An approach that circumvents this difficulty is presented in Chapter 4 and is summarized under Item 3 below;

2. Capture heterogeneity of network users (travelers) with regard to their willingness to pay as it is reflected in the relative valuation of the attributes affecting route selection, as well as other choice dimensions. As discussed, this could be accomplished in an approximate manner through segmentation of the user population and solution of a multiclass traveler assignment model with predefined classes, although it is shown in Chapter 4 that an approach based on a continuous distribution of VOT (or other choice attribute coefficient) is preferable and could be implemented in connection with any network-simulation–based network-loading procedure. Such an approach would require implementation of a new parametric path-finding procedure, as well as relatively minor modification of the loading and updating process (from one iteration to the next). This procedure was successfully demonstrated for the New York best practice model network;
3. Generate path-level reliability measures in order to incorporate reliability as an attribute in path choice decisions (and other choice dimensions as applicable). This is perhaps the most challenging of the elements of the methodology developed to address the objectives of the study. For this reason, it is the subject of a separate research study under a SHRP 2 Reliability program (Project L04). In that project, the goal is to generate measures of reliability from micro, meso, and possibly macro traffic-simulation tools. In this work, a relatively simple approach was developed for planning applications that relies on a robust relation between the mean travel time per unit distance and the corresponding standard deviation of the mean travel time, also per unit distance. The normalization by the traveled distance is essential for the relation to hold. This relation can be applied directly at the route or origin–destination (O-D) level, thereby circumventing the nonadditivity issue noted in connection with Item 1 above. This relation builds on classic work in traffic science and has been extensively validated for simulated network results. In addition, it is currently being further tested using trajectory data from mobile vehicle probes in connection with SHRP 2 Project L04. Chapter 4 shows how this relation is applied directly at the path level to estimate reliability (in the form of standard deviation given the mean value). This approach could be readily implemented with any simulation tool; and
4. Cope with large-scale network issues by reducing computational requirements. As simulation-based assignment methods, which are an essential platform for capturing user responses to pricing and reliability, are applied to realistic regional networks comparable to those typically used for static assignment applications, it is essential that their computational requirements become manageable. Simple, single-class applications with homogeneous users have long been executable on very large networks without much difficulty. Addressing the objectives of the present study required pushing the frontier with regard to the network size that may be executed with the advanced tools for heterogeneous users. Through several computer science and algorithmic implementation techniques, it was possible to reduce the computational requirements for very large networks (as shown with the New York data) and successfully demonstrate the computational reductions. These reductions are associated primarily with the path-finding and equilibration algorithms developed for multiuser classes. Although additional improvement is undoubtedly possible in this regard, the present implementation to New York demonstrates the feasibility of such a large-scale application and paves the way for bigger and faster procedures in the future.

All the methods proposed in this study have been demonstrated in connection with a state-of-the-art simulation-based DTA methodology. The DYNASMART-P code provided the implementation platform, although the components of the implementation and modifications could be replicated in a straightforward manner in connection with any simulation-based DTA tool. In particular, these methods could be directly implemented in mesosimulation DTA models such as Dynamit-P, which follows a similar blueprint; Dynus-T, which is built directly on the basic DYNASMART-P platform; VISTA, which adapts a similar network representation and path-finding algorithmic structure in a modified simulation platform; and several other meso-DTA tools. Because these procedures apply primarily to the path-finding procedures and are independent of the simulation logic, they can be readily implemented with microlevel simulators, although the latter cannot at this stage address large-scale networks of the scale of the New York City regional network. This drawback has prompted several microscopic simulation software vendors to release mesoscopic versions of their tools (e.g., Aimsun). Application in TRANSIMS is also possible, though it requires more effort because of its elaborate router, which does not entail traditional path finding and would therefore require adaptation to its shortest-path computation method.

Although the use of static assignment tools is not recommended in conjunction with studies of pricing schemes and reliability assessment, it is nonetheless possible to incorporate

some of the above methods and recommendations in static assignment tools. In particular, if path-based assignment approaches are used, it is possible to improve the route choice basis by implementing the recommended options. For instance, shortest-path-finding procedures can readily consider multiple attributes in a generalized cost function when these attributes are additive across links. Incorporating heterogeneity in user preferences can also be accomplished by using a parametric shortest-path method, originally proposed by Leurent (1993) and Dial (1997). Incorporating travel time variability is also possible, at least in aggregate form, using the same robust relation already discussed. With path-based assignment implementations, the structural differences between static and dynamic assignment become less pronounced, and it is possible to essentially adapt all the discussion and methods to encompass static assignment techniques, although the critical time dimension of the problem would be inherently ignored.

Integrated Demand and Network Models

Two-Way Linkage Between Travel Demand and Network Supply

Since the technologies of microsimulation have been brought to a certain level of maturity on both the demand side (ABM) and the supply (network) side (DTA), the challenge of ABM-DTA integration has become one of the most promising avenues in transportation modeling. Seemingly, the integration between the two models should be as natural and straightforward as the integration concept between a four-step model and static traffic assignment, shown in Figure 5.2. The relatively simple integration of the demand and supply sides in the conventional framework is based on the fact that the input and output entities involved in the process have the same matrix structure. The four-step demand model produces trip tables needed for assignment, and the assignment procedures produce full LOS skims in the same matrix format needed for the four-step model. Note that the LOS variables are provided for all possible trips (not only for the trips generated by the demand model at the current iteration). In this case it can be said that the network model provides a full feedback to the demand model. The theory of global demand-network equilibrium is well developed for this case and guarantees a unique solution for the problem, as well as a basis for effective practical algorithms.

Both ABM and DTA, however, operate with individual particles as modeling units (individual tours and trips) and can have compatible levels of spatial and temporal resolution. It might seem that exactly the same integration concept as applied for four-step models could be adjusted to account for a list of individual trips instead of fractional-number trip tables.

Moreover, the advanced individual ABM-DTA framework would provide an additional beneficial dimension for the integration in the form of consistent individual schedules, which can never be incorporated in an aggregate framework. Individual schedule consistency means that for each person, the daily schedule (i.e., a sequence of trips and activities) is formed without gaps or overlaps.

However, a closer look at the ABM-DTA framework and consideration of the actual technical aspects of implementation reveal some nontrivial issues that need to be resolved before the advantages offered by overall microsimulation framework can be taken. The specific problem is illustrated in Figure 5.3, which shows that the feedback provided by the DTA procedure does not cover all the needs of the ABM.

The crux of the problem is that unlike the four step–static traffic assignment integration, the microsimulation DTA can only produce an individual trajectory (path in time and space) for the list of actually simulated trips. It does not automatically produce trajectories for all (potential) trips to other destinations and at other departure times. Thus, it does not provide the necessary LOS feedback to ABM at the disaggregate level for all modeled choices. Any attempt to resolve this issue by "brute force" would result in an infeasibly large number of calculations, since all possible trips cannot be processed by

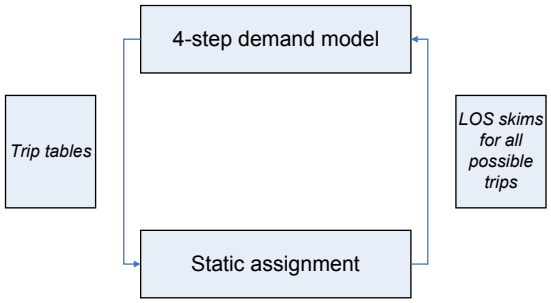

Figure 5.2. Integration of four-step model and static assignment.

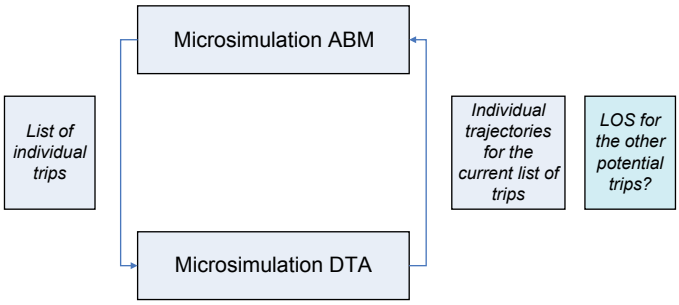

Figure 5.3. Integration of ABM and DTA (direct).

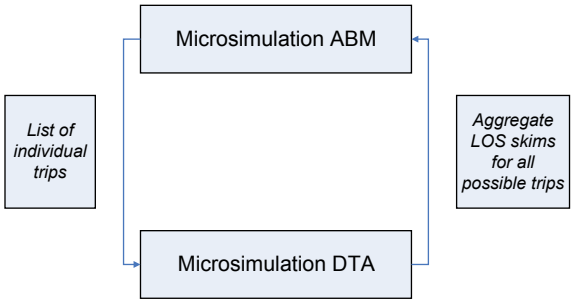

Figure 5.4. Integration of ABM and DTA (aggregate feedback).

DTA at the disaggregate level. In fact, the list of trips for which the individual trajectories can be produced is a very small portion of the all possible trips to consider.

As shown in Figure 5.4, one of the possible solutions is to employ DTA to produce crude LOS matrices (the way they are produced by static traffic assignment) and use these LOS variables to feed the demand model. However, this approach, in the aggregation of individual trajectories into crude LOS skims, would lose most of the detail associated with DTA and the advantages of individual microsimulation (e.g., individual variation in VOT or other person characteristics). Essentially, with this approach the individual schedule consistency concept would be of limited value because travel times would be crude for each particular individual. Nevertheless, this approach has been adopted in many studies due to its inherent simplicity (Bekhor et al. 2011; Castiglione et al. 2012).

Instead of the model integration ideas outlined above, the team proposes several new ideas that are currently being considered and tested in the SHRP 2 L04 project. These ideas are explained in the subsequent sections.

Activity-Based Model–Dynamic Traffic Assignment Integration Principles

The emphasis in the L04 project is on truly integrating the demand and network models, not merely connecting them through aggregate measures in an iterative application. The team's approach is based on the following principles:

- Use of a fully disaggregate approach implemented at the disaggregate individual level (travel tours by person);
- Conceptual integration of the demand and network simulation procedures to ensure a fully consistent daily schedule for each individual. This approach is principally different from the so-called "iterative loose coupling" of the demand and supply models. The basic travel unit exchanged between ABM and DTA is a travel tour, rather than an elemental trip;
- Representation of user heterogeneity (individual travel variations) in network-based choice processes, with implications for optimum-path computations;
- Use of new algorithms that fully exploit the particle-based (individual) representation of vehicles flowing through the network in computing equilibria or other demand–supply consistent states;
- Recognition that different policies call for different types of solutions, with varying degrees of user information and feedback, such as nonrecurrent congestion with limited or local information that would call for one-shot simulations, versus recurrent congestion that calls for a long-term dynamic equilibrium solution, versus applications in which day-to-day learning and evolution may be more important than the final states; and
- Use of advanced concepts from agent-based modeling for integrating behavior processes in a network context, with special-purpose data structures geared to the physical and behavioral processes modeled.

Consistency of Individual Daily Schedule

The concept of a fully consistent individual daily schedule is illustrated in Table 5.1. The daily schedule of a person is modeled for 24 hours starting at 3:00 a.m. on the simulation day and ending at 3:00 a.m. next day (formally represented as 27:00). The integrated model operates with four schedule-related types of events: (1) in-home activities, (2) out-of-home activities, (3) trips, and (4) tours. Start and end times of activities logically relate to the corresponding departure and arrival times of trips connecting these activities. Each tour spans several trips and related out-of-home activities and essentially represents a fragment of the individual daily schedule.

In reality, the observed individual schedules are always consistent in the sense that they obey time–space constraints and have a logical, continuous timeline in which all activities and trips are sequenced with no gaps and no overlaps. However, achieving full consistency has not yet been resolved in operational models. The crux of the problem is that all trips and associated activities have to obey a set of hard (physical) and soft (consideration of probabilistic choices) constraints that can only partially be taken into account without a full integration between the demand and network simulation models. Also, both models should be brought to a level of temporal resolution that is sufficient for controlling the constraints (e.g., 5 minutes).

The following constraints should be taken into account:

- Schedule Continuity. Activity start time should correspond to the preceding trip arrival time, and activity end time should correspond to the following trip departure time. This hard constraint is not controlled in either the four-step

Table 5.1. Fully Consistent Individual Daily Schedule

In-home			Trips			Out-of-Home			Tours		
Activity	**Start**	**End**	**Purpose**	**Depart**	**Arrive**	**Activity**	**Start**	**End**	**Purpose**	**Depart**	**Arrive**
Sleeping, eating at home, errands	3:00										
		7:30	Escort	7:30					Work	7:30	
					7:45	Drop off child at school	7:45				
			Work	7:50				7:50			
					8:30	Work	8:30				
			Shop	16:30				16:30			
					17:00	Shop	17:00				
			Return home	17:30				17:30			
Child care, errands	18:00				18:00						18:00
		19:00	Disc	19:00					Disc	19:00	
					19:30	Theater	19:30				
			Return home	21:30				21:30			
Resting, errands, sleeping	22:00				22:00						22:00
		27:00									

demand models or the static trip-based network simulation models because they operate with unconnected trips and do not control for activity durations at all. Also, in four-step models, the inherently crude level of temporal resolution does not allow for incorporating this constraint. In ABMs, starting from the Columbus model developed in 2004, certain steps have been made to ensure a partial consistency between departure and arrival times, as well as duration at the entire-tour level (Vovsha and Bradley 2004). However, these improvements do not include trip details, and they do not control for feasibility of travel times within the tour framework (although travel time is used as one of the explanatory variables). Certain attempts to incorporate trip departure-time choice in a framework of trip chains have been made within DTA models (Abdelghany and Mahmassani 2003). However, these attempts were limited to a tour level only, and they also required a simplified representation of activity duration profiles. This constraint expresses consistency (i.e., the same number) in each row of Table 5.1.

- Physical Flow Process Properties. These hard constraints apply to network loading and flow propagation aspects in DTA procedures. Physical principles such as conservation of vehicles at nodes must be adhered to strictly (e.g., no vehicles should simply be lost or otherwise disappear from the system). This constraint accounts for feasibility of travel times obtained in the network simulation that are further used to determine trip departure and arrival times in the corresponding columns of Table 5.1.
- Equilibrium Travel Times. Travel times between activities in the schedule generated by the demand model should correspond to realistic network travel times for the corresponding origin, destination, departure time, and route generated by the traffic simulation model with the given demand. While most of the four-step models and ABMs include a certain level of demand–supply equilibration, they are limited to achieving stability in terms of average travel times. There is no control for consistency within the individual daily schedule. The challenge is to couple this constraint with the previous one; that is, to ensure individual schedule continuity with equilibrium travel times. This hard constraint expresses consistency between trip departure and arrival times in the corresponding columns of Table 5.1 with the travel times obtained in the network simulation. Practically, it is achieved within a certain tolerance level.
- Realistic Activity Timing and Duration. Activities in the daily schedule have to be placed according to behaviorally realistic temporal profiles. Each activity has a preferred start time, end time, and duration formalized as a utility function with multiple components. In the presence of congestion and pricing, travelers may deviate from the preferred temporal profiles (as well as even cancel or change the order of activity episodes). However, this rescheduling process should

obey utility-maximization rules over the entire schedule and cannot be effectively modeled by simplified procedures that adjust departure times for each trip separately. None of the existing operational ABMs explicitly controls for activity durations, although some of them control for entire-tour durations (such as the MTC ABM) or the duration of the activity at the primary destination (as implemented in the SACOG ABM). DTA models that incorporate departure time choice have been bound to a simplified representation of temporal utilities and limited to trip chains in order to operate within a feasible dimensionality of the associated choices when combined with the dynamic route choice. This soft constraint expresses consistency between activity start and end times in the corresponding columns of Table 5.1 with the schedule utility-maximization principle (or in a more general sense, with the observed timing and duration pattern for activity participation). In operational models, the focus has been primarily on out-of-home activities. It should be noted, however, that it is also important to preserve a consistent and realistic pattern of in-home activities (e.g., reasonable time constraints for sleeping and household activities), as well as take into account possible substitution between in-home and out-of-home durations for work, shopping, and discretionary activities.

Schedule consistency with respect to all five constraints is absolutely essential for time-sensitive policies like congestion pricing. In reality, any change in the timing of a particular activity or trip spurred by a congestion pricing policy would trigger a chain of subsequent adjustments through the whole individual schedule. It can be shown that under certain circumstances, an attempt to alleviate congestion in the a.m. period by pricing may result in worsening congestion in the p.m. period because of the compression of individual daily schedules that are forced to start later (PB Consult, Inc. 2005).

In order to address all five constraints, the model system has to be truly integrated with a mutual core between the ABM and DTA modules. This mutual core has to fully address the temporal dimension of activities and trips, but other choice dimensions can be effectively treated by each corresponding module, as shown in Figure 5.5.

The mutual core ensures synchronization of time-related ABM and DTA components and is designed to achieve a full schedule consistency at the individual level. The ABM model generates tours with origins, destinations, and trip departure

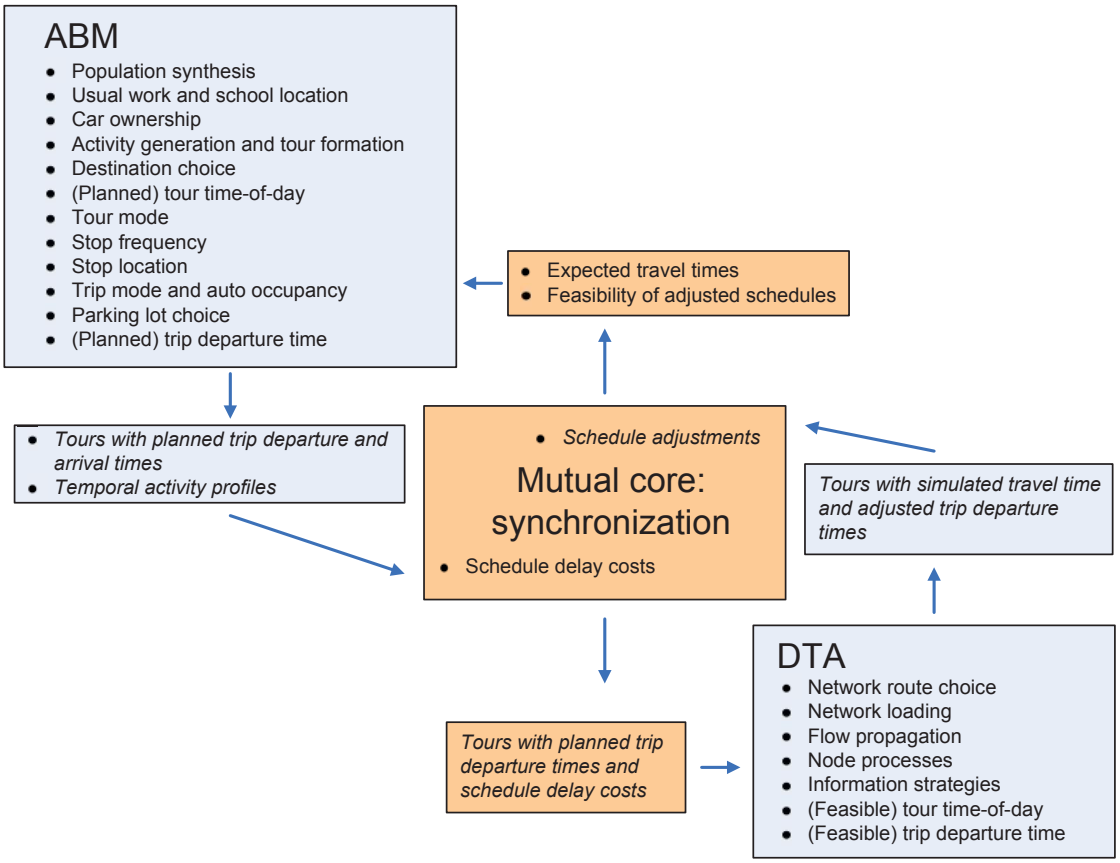

Figure 5.5. Integration scheme of ABM and DTA.

times based on expected travel times (from the DTA) and TOD choice utilities. These can be converted to temporal activity profiles for each activity episode; the temporal activity profile is essentially an expected utility of activity participation for a given time unit. As discussed above, these temporal activity profiles can be converted into schedule delay cost functions for each trip arrival time, which are input to the DTA model.

The DTA model assigns each trip on the network, determines the route, and reschedules trip departure times based on the feasible travel times (which may be different from the expected travel times used in the ABM). This rescheduling is done based on the updated congested travel times; it takes into account schedule delay cost, as well as interdependencies across trips on the same tour. These features have been recently added to the DTA algorithm and tested for DYNASMART-P (Abdelghany and Mahmassani 2003; Zhou et al. 2008). The capability of DTA to handle travel tours rather than trips is essential to ensure consistency between DTA and ABM. Individual choices are to be resimulated even if the DTA was able to fulfill the planned schedule successfully. For subsequent iterations, after aggregate travel times have been stabilized, a (gradually diminishing) portion of individuals will be subject to demand resimulation, and these individuals will be chosen on the basis of the feasibility of their adjusted schedules and the magnitude of the adjustments introduced by DTA. The team's research on equilibration of the integrated models has resulted in new procedures for directing the convergence algorithm toward a mutually consistent solution through selection of the fraction of individuals or households whose schedules may be replanned in each iteration.

After each tour has been adjusted, the synchronization module consolidates the entire daily schedule for each individual. Depending on the magnitude of adjustments, the schedule might result in an infeasible (or highly improbable) state in which tours are overlapped or activity durations have reached unreasonable values. The synchronization module informs the ABM which individual daily schedules have to be resimulated. Individuals whose schedules have to be resimulated will undergo a complete chain of demand choices based on the updated travel times.

Individual Schedule Adjustments (Temporal Equilibrium)

Integration of ABM and DTA at a disaggregate level of individual trips requires an additional model component to be developed. This component acts as an interface that transforms the DTA output (individual vehicle trajectories), with departure and arrival times for each trip simulated with a high level of temporal resolution, into schedule adjustments to the individual schedules generated by the ABM. The purpose of this feedback is to achieve consistency between generated activity schedules (activity start times, end times, and durations) and trip trajectories (trip departure time, duration, and arrival time). This feedback is implemented as part of the temporal equilibrium between ABM and DTA, when all trip destinations and modes are fixed, but departure times are adjusted until a consistent schedule is built for each individual.

Individual schedule consistency means that for each person, the daily schedule (i.e., a sequence of trips and activities) is formed without gaps or overlaps, as shown in Figure 5.6. In this way, any change in travel time would affect activity durations and vice versa.

New methods for equilibrating ABM and DTA are presented in Figure 5.7, in which two innovative technical solutions are applied in parallel. The first solution is based on the fact that a direct integration at the disaggregate level is possible along the temporal dimension if the other dimensions (number of trips, order of trips, and trip destinations) are fixed for each individual. Full advantage can then be taken of the individual schedule constraints and corresponding effects, as shown in Figure 5.6. The inner loop of temporal equilibrium includes schedule adjustments in individual daily activity patterns that occur when congested travel times

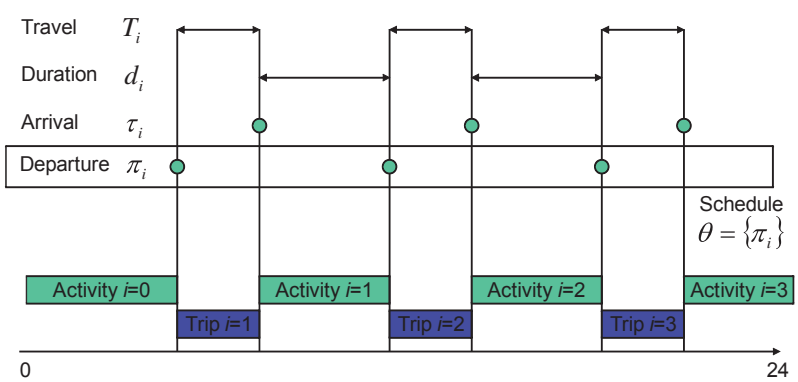

Figure 5.6. Consistent individual daily schedule.

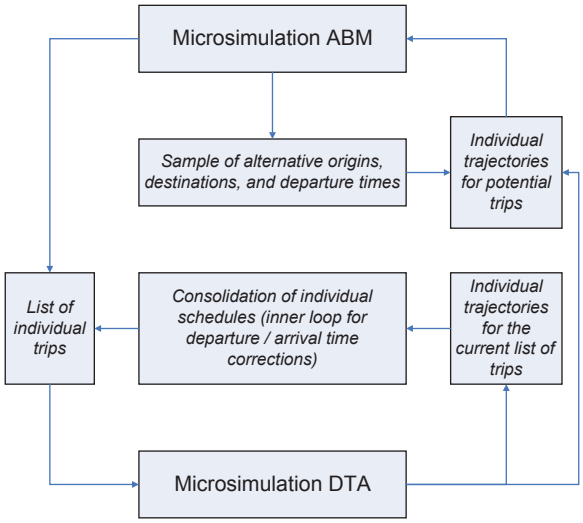

Figure 5.7. Integration of ABM and DTA (split feedback).

are different from planned travel times. This action helps the DTA to reach convergence (inner loop), and is nested within the global system loop (when the entire ABM is rerun and demand is regenerated).

The second solution is based on the fact that trip origins, destinations, and departure times can be presampled, and the DTA process would only be required to produce trajectories for a subset of origins, destinations, and departure times. In this case, the schedule consolidation is implemented though corrections of the departure and arrival times (based on the individually simulated travel times) and is employed as an inner loop. The outer loop includes a full regeneration of daily activity patterns and schedules, but with a subsample of locations for which trajectories are available (it also can be interpreted as a learning and adaptation process with limited information).

Adjustment of the individual daily schedule can be formulated as an entropy-maximizing problem of the following form:

$$\min_{\{x_i\}} \left\{ \begin{array}{l} \left[\sum_{i=0}^{I} w_i \times x_i \times \ln\left(\frac{x_i}{d_i}\right) \right] + \left[\sum_{i=1}^{I+1} u_i \times y_i \times \ln\left(\frac{y_i}{\pi_i}\right) \right] \\ + \left[\sum_{i=0}^{I} v_i \times z_i \times \ln\left(\frac{z_i}{\tau_i}\right) \right] \end{array} \right\} \tag{5.1}$$

subject to

$$y_i = \tau_0 + \left(\sum_{j=0}^{i-1} x_j \right) + \left(\sum_{j=0}^{i-1} t_j \right), \qquad i = 1, 2, \ldots, I+1 \tag{5.2}$$

$$= \tau_0 + \left(\sum_{j=0}^{i-1} x_j \right) + \left(\sum_{j=0}^{i} t_j \right), \qquad i = 1, 2, \ldots, I \tag{5.3}$$

$$x_i > 0, \qquad i = 0, 1, 2, \ldots, I \tag{5.4}$$

where

$i, j = 1, 2, \ldots, I$ = trips and associated activities at the trip destination;
$i, j = 0$ = activity at home before the first trip;
$i = I + 1$ = (symbolic) departure from home at the end of the simulation period;
x_i, x_j = adjusted activity duration;
y_i = adjusted departure time for trip to the activity;
z_i = adjusted arrival time for trip to the activity;
d_i = planned activity duration;
π_i = planned departure time for trip to the activity;
τ_i = planned arrival time for trip to the activity;
t_i, t_j = actual time for trip to the activity that is different from expected;
w_i = schedule weight (priority) for activity duration;
u_i = schedule weight (priority) for trip departure time; and
v_i = schedule weight (priority) for trip arrival time.

The essence of this formulation is that in the presence of travel times that are different from the expected travel times that the user used to build the schedule, it will try to accommodate new travel times in such a way that the schedule is preserved to the extent possible. This preservation relates to activity start times (trip arrival times), activity end times (trip departure times), and activity durations (Equation 5.1). The relative weights represent the priorities of different activities in terms of start time, end time, and duration. The greater the weight, the more important it is for the user to keep the corresponding component close to the original schedule. Very large weights correspond to inflexible, fixed-time activities. The weights directly relate to the schedule delay penalties as described below in the section on travel time reliability measures. The concept of schedule delay penalties relates to deviation from the (preferred or planned) activity start time (trip arrival time) only, but the schedule adjustment formulation allows for a joint treatment of deviations from the planned start times, end times, and durations.

The constraints express the schedule consistency rule as shown in Figure 5.6. Equation 5.2 expresses departure time for each trip as a sum of the previous activity durations and travel times. Equation 5.3 expresses arrival time of each trip as a sum of the previous activity durations and travel times plus travel time for the given trip. The (symbolic) arrival time for the home activity prior to the first trip is used to set

the scale of all departure and arrival times. In this way, the problem is formulated in the space of activity durations, and the trip departure and arrival times are derived from the activity durations and given travel times.

The solution of the convex problem can be found by writing the Lagrangian function and equating its partial derivatives (with respect to activity durations) to zero. The equation takes the following form:

$$x_i = d_i \times \left\{ \prod_{j>1} \left[\left(\frac{\pi_j}{y_j} \right)^{u_j} \times \left(\frac{\tau_j}{z_j} \right)^{v_j} \right] \right\}^{\frac{1}{w_i}} \tag{5.5}$$

This solution is easy to find by using either an iterative balancing method or the Newton–Raphson method. The essence of this formula is that updated activity durations are proportional to the planned durations and adjustment factors. The adjustment factors are applied considering the duration priority. If the duration weight is very large, then the adjustments will be minimal. The duration adjustment is calculated as a product of trip departure and arrival adjustments for all subsequent trips. The trip departure adjustment $\left\{ \frac{\pi_j}{y_j} \right\}$ and trip arrival adjustment $\left\{ \frac{\tau_j}{z_j} \right\}$ can be interpreted as lateness versus the planned schedule if it is less than one and earliness if it is greater than one. Each trip departure or arrival adjustment factor is powered by the corresponding priority weight. As a result, activity duration will shrink if there are many subsequent trip departures or arrivals (or both) that are later than planned. Conversely, activity duration will be stretched if there are many subsequent trip departures or arrivals (or both) that are earlier than planned. Overall, the model seeks the equilibrium (compromise) state in which all activity durations, trip departures, and trip arrivals will be adjusted to accommodate the changed travel times while preserving the planned schedule components by priority.

The team provides demonstration software and has implemented many numerical tests with this model. In particular, the iterative balancing procedure goes through the following steps:

1. Set initial activity durations equal to the planned durations $\{x_i = d_i\}$;
2. Update trip departure times with new travel times and updated activity durations using Equation 5.2;
3. Update trip arrival times with new travel times and updated activity durations using Equation 5.3;
4. Calculate balancing factors $\left\{ \frac{\pi_j}{y_j} \right\}$ for trip departure times (lateness if less than one, earliness if greater than one);
5. Calculate balancing factors $\left\{ \frac{\tau_j}{z_j} \right\}$ for trip arrival times (lateness if less than one, earliness if greater than one);
6. Update activity durations using Equation 5.5; and
7. Check for convergence with respect to activity durations; if not, go to Step 2.

Applying this model in practice requires default values for activity durations, trip departure times, and trip arrival times. This is an area for which more specific data on schedule priorities and constraints of different person types would be welcomed. This type of data is already included in some household travel surveys with respect to work schedules. It should be extended to include nonwork activities, many of which also have schedule constraints. At this stage, the team suggests the default values shown in Table 5.2.

If some activity in the schedule falls into more than one category (e.g., work and first activity of the day), the maximum weight is applied from each column.

Incorporation of Reliability in Demand Model

The proposed methods of quantification of reliability should be incorporated in the demand model (ABM) with respect to subchoices such as tour and trip mode choice, destination choice, and TOD choice. In the typical ABM structure, a generalized cost function with a reliability term can be directly included in the utility function for highway modes. Further on, the reliability term will affect destination and TOD choice through mode choice logsums. In the same vein, it affects upper-level choice models of car ownership and activity–travel patterns though accessibility measures that represent simplified destination choice logsums. As discussed in Chapter 3, the demand side of travel time reliability has been explored in detail in the SHRP 2 C04 project. In this section, the team presents a concise overview of each method and its applicability in an operational travel demand model.

Perceived Highway Time in Demand Model

This method is easy to implement without a significant restructuring of the demand model. Essentially, the generic highway travel time variable in mode choice should be replaced with segmented travel time by congestion levels using the weights recommended in Table 5.3.

The weights applied have to be consistent between traffic assignment and mode choice. The table provides pivot points that can be interpolated between them linearly using the volume-to-capacity ratio or flow density parameter. However, perceived travel time is not a direct measure of travel time

Table 5.2. Recommended Weights for Schedule Adjustment

Activity Type	Duration	Trip Departure to Activity	Trip Arrival at Activity Location
Work (low income)	5	1	20
Work (high income)	5	1	5
School	20	1	20
Last trip to activity at home	1	1	3
Trip after work to NHB activity	1	5	1
Trip after work to NHB activity	1	10	1
NHB activity on at-work subtour	1	5	5
Medical	5	1	20
Escorting	1	1	20
Joint discretionary, visiting, eating out	5	5	10
Joint shopping	3	3	5
Any first activity of the day	1	5	1
Other activities	1	1	1

Note: NHB = nonhome based.

reliability. It can be used as a surrogate when more advanced methods are not available, but it is less appealing behaviorally and it is not the main focus of the current research.

Mean Variance in Demand Model

This method is easy to implement and does not require a significant restructuring of the demand model. Essentially, it requires an inclusion of an additional reliability term in the mode choice utility for highway modes. The following form of generalized cost component in the mode utility function can be recommended as the first step for incorporation in operational models (many additional modifications and nonlinear transformations are analyzed in Chapter 3):

$$U = a \times T + b \times C + c \times SD(T) \tag{5.6}$$

Table 5.3. Recommended Highway Travel Time Weight by Congestion Levels

Travel Time Condition	Weight
Free flow	1.00
Busy	1.05
Light congestion	1.10
Heavy congestion	1.20
Stop start	1.40
Gridlock	1.80

where

T = mean travel time;
C = travel cost;
$SD(T)$ = standard deviation of travel time;
a = coefficient for travel time;
b = coefficient for travel cost;
c = coefficient for standard deviation of travel time;
a/b = VOT;
c/b = value of reliability (VOR); and
c/a = reliability ratio (ρ = VOT/VOR).

Recommended values for the parameters are summarized in Table 5.4. The parameters are segmented by travel purpose, household income, car occupancy, and travel distance.

Schedule Delay Cost in Demand Model

Several models were estimated in the SHRP 2 C04 research with schedule delay cost as described in Chapter 3. The majority of them were estimated using different stated preference settings in which either route or departure time served as the underlying travel choice dimension. The technical details for the inclusion of this method in an operational travel demand model have not yet been fully explored. As shown in Figure 5.8, the team outlines two possible approaches that differ in how and where the schedule delay cost component is calculated.

In both approaches, the travel demand model (its TOD choice or activity scheduling submodel) produces a preferred departure time (PDT) and preferred arrival time (PAT) for

Table 5.4. Recommended Values of Parameters for Generalized Cost Function with Reliability

Travel Purpose	Examples of Population and Travel Characteristics			Model Coefficients and Derived Measures					
	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Cost Coefficient with Income and Occupancy Effects	Cost for SD (min)	VOT ($/h)	VOR ($/h)	Reliability Ratio
Work and Business	30,000	1.0	5.0	–0.0425	–0.0026	–0.1042	9.9	24.3	2.45
	30,000	2.0	5.0	–0.0425	–0.0015	–0.1042	17.2	42.3	2.45
	30,000	3.0	5.0	–0.0425	–0.0011	–0.1042	23.9	58.5	2.45
	30,000	1.0	10.0	–0.0425	–0.0026	–0.0521	9.9	12.1	1.23
	30,000	2.0	10.0	–0.0425	–0.0015	–0.0521	17.2	21.1	1.23
	30,000	3.0	10.0	–0.0425	–0.0011	–0.0521	23.9	29.2	1.23
	30,000	1.0	20.0	–0.0425	–0.0026	–0.0260	9.9	6.1	0.61
	30,000	2.0	20.0	–0.0425	–0.0015	–0.0260	17.2	10.6	0.61
	30,000	3.0	20.0	–0.0425	–0.0011	–0.0260	23.9	14.6	0.61
	60,000	1.0	5.0	–0.0425	–0.0017	–0.1042	15.0	36.8	2.45
	60,000	2.0	5.0	–0.0425	–0.0010	–0.1042	26.1	64.1	2.45
	60,000	3.0	5.0	–0.0425	–0.0007	–0.1042	36.2	88.6	2.45
	60,000	1.0	10.0	–0.0425	–0.0017	–0.0521	15.0	18.4	1.23
	60,000	2.0	10.0	–0.0425	–0.0010	–0.0521	26.1	32.0	1.23
	60,000	3.0	10.0	–0.0425	–0.0007	–0.0521	36.2	44.3	1.23
	60,000	1.0	20.0	–0.0425	–0.0017	–0.0260	15.0	9.2	0.61
	60,000	2.0	20.0	–0.0425	–0.0010	–0.0260	26.1	16.0	0.61
	60,000	3.0	20.0	–0.0425	–0.0007	–0.0260	36.2	22.2	0.61
	100,000	1.0	5.0	–0.0425	–0.0013	–0.1042	20.4	50.0	2.45
	100,000	2.0	5.0	–0.0425	–0.0007	–0.1042	35.5	87.1	2.45
	100,000	3.0	5.0	–0.0425	–0.0005	–0.1042	49.1	120.4	2.45
	100,000	1.0	10.0	–0.0425	–0.0013	–0.0521	20.4	25.0	1.23
	100,000	2.0	10.0	–0.0425	–0.0007	–0.0521	35.5	43.5	1.23
	100,000	3.0	10.0	–0.0425	–0.0005	–0.0521	49.1	60.2	1.23
	100,000	1.0	20.0	–0.0425	–0.0013	–0.0260	20.4	12.5	0.61
	100,000	2.0	20.0	–0.0425	–0.0007	–0.0260	35.5	21.8	0.61
	100,000	3.0	20.0	–0.0425	–0.0005	–0.0260	49.1	30.1	0.61

(continued on next page)

Table 5.4. Recommended Values of Parameters for Generalized Cost Function with Reliability (continued)

Travel Purpose	Examples of Population and Travel Characteristics			Model Coefficients and Derived Measures					
	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Cost Coefficient with Income and Occupancy Effects	Cost for SD (min)	VOT ($/h)	VOR ($/h)	Reliability Ratio
Nonwork	30,000	1.0	5.0	–0.0335	–0.0030	–0.0697	6.7	13.8	2.08
	30,000	2.0	5.0	–0.0335	–0.0019	–0.0697	10.8	22.5	2.08
	30,000	3.0	5.0	–0.0335	–0.0014	–0.0697	14.4	29.9	2.08
	30,000	1.0	10.0	–0.0335	–0.0030	–0.0348	6.7	6.9	1.04
	30,000	2.0	10.0	–0.0335	–0.0019	–0.0348	10.8	11.2	1.04
	30,000	3.0	10.0	–0.0335	–0.0014	–0.0348	14.4	14.9	1.04
	30,000	1.0	20.0	–0.0335	–0.0030	–0.0174	6.7	3.5	0.52
	30,000	2.0	20.0	–0.0335	–0.0019	–0.0174	10.8	5.6	0.52
	30,000	3.0	20.0	–0.0335	–0.0014	–0.0174	14.4	7.5	0.52
	60,000	1.0	5.0	–0.0335	–0.0021	–0.0697	9.4	19.6	2.08
	60,000	2.0	5.0	–0.0335	–0.0013	–0.0697	15.3	31.8	2.08
	60,000	3.0	5.0	–0.0335	–0.0010	–0.0697	20.3	42.3	2.08
	60,000	1.0	10.0	–0.0335	–0.0021	–0.0348	9.4	9.8	1.04
	60,000	2.0	10.0	–0.0335	–0.0013	–0.0348	15.3	15.9	1.04
	60,000	3.0	10.0	–0.0335	–0.0010	–0.0348	20.3	21.1	1.04
	60,000	1.0	20.0	–0.0335	–0.0021	–0.0174	9.4	4.9	0.52
	60,000	2.0	20.0	–0.0335	–0.0013	–0.0174	15.3	8.0	0.52
	60,000	3.0	20.0	–0.0335	–0.0010	–0.0174	20.3	10.6	0.52
	100,000	1.0	5.0	–0.0335	–0.0017	–0.0697	12.2	25.3	2.08
	100,000	2.0	5.0	–0.0335	–0.0010	–0.0697	19.8	41.1	2.08
	100,000	3.0	5.0	–0.0335	–0.0008	–0.0697	26.2	54.6	2.08
	100,000	1.0	10.0	–0.0335	–0.0017	–0.0348	12.2	12.6	1.04
	100,000	2.0	10.0	–0.0335	–0.0010	–0.0348	19.8	20.5	1.04
	100,000	3.0	10.0	–0.0335	–0.0008	–0.0348	26.2	27.3	1.04
	100,000	1.0	20.0	–0.0335	–0.0017	–0.0174	12.2	6.3	0.52
	100,000	2.0	20.0	–0.0335	–0.0010	–0.0174	19.8	10.3	0.52
	100,000	3.0	20.0	–0.0335	–0.0008	-0.0174	26.2	13.6	0.52

Note: SD = standard deviation.

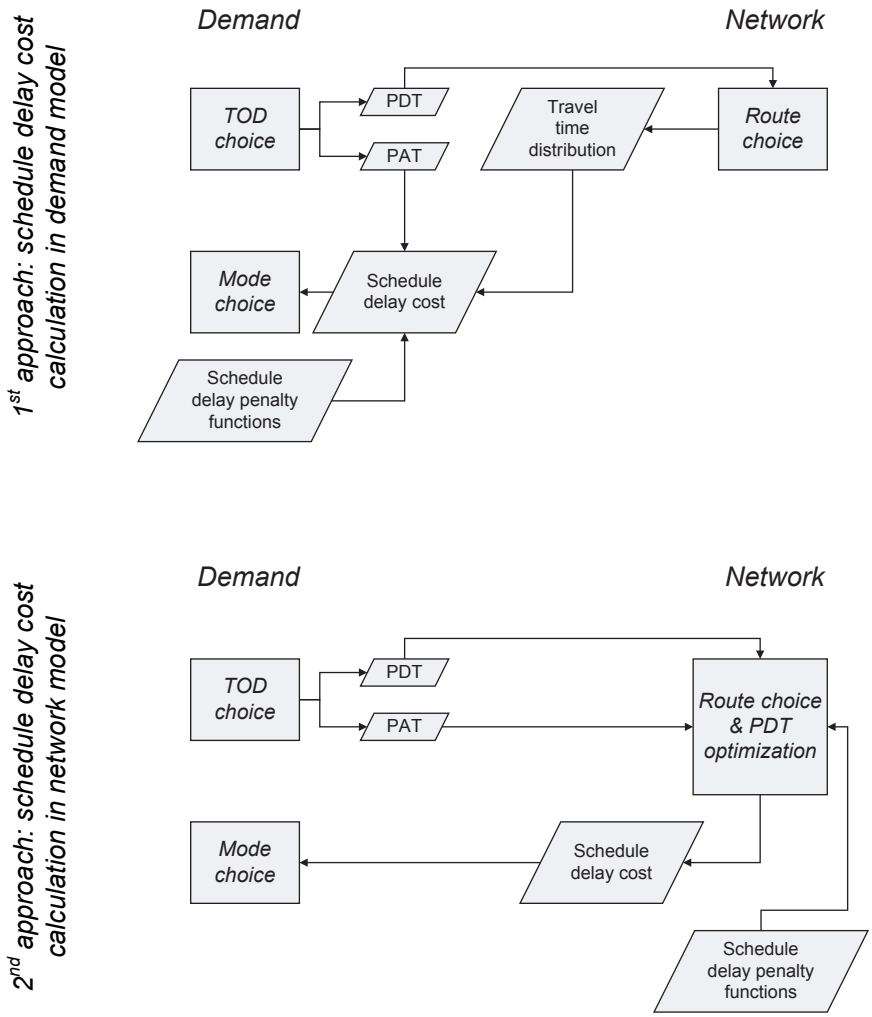

Figure 5.8. Incorporation of schedule delay cost calculation in* (top) *demand model and* (bottom) *network model (mode choice).

each trip based on the expected travel times (and known variations if used in the scheduling procedure and departure time optimization). In both approaches, schedule delay penalty functions are assumed known for each trip. The principal difference is in how the demand model interacts with the network simulation model to produce the expected scheduled delay cost for each trip.

In the first approach, schedule delay cost is calculated in the demand model as part of the mode utility calculation for highway modes. The network simulation model assigns trips based on PDT without considering PAT. The role of the network simulation model is to produce travel time distributions for each trip (through a single equilibrium run or multiple runs). Subsequently, schedule delay cost is integrated over the travel time distribution in the demand model. This scheme has not yet been tested. The most realistic implementation approach for this scheme is a multiple-run framework, which is discussed below.

In the second approach, the calculation of schedule delay cost is incorporated in the network model and is fed into the demand model. Perhaps the most behaviorally appealing aspect of this implementation approach occurs when the network simulation model is allowed to optimize PDT based on PAT and specified schedule delay penalties. This means that the route choice component is replaced with a joint route and departure time choice. This type of model can be implemented in a single-run framework; some testing of this approach has been reported (Zhou et al. 2008).

In both cases, the main (technical) obstacle for practical implementation of the schedule delay approach is the necessity to generate PAT for each trip against which the schedule delay cost is calculated as a consequence of unreliable travel time. It is currently unrealistic to prepare PAT as an input to travel demand models, although for some trips with inherently fixed schedules (e.g., work with fixed schedule, appointments, ticketed shows), this might be ultimately the right

approach. Some approaches to endogenously calculated PAT within the scheduling model as a latent variable were suggested by Ben-Akiva and Abou-Zeid (2007). Further research is needed to operationalize this approach within the framework of a regional travel model.

Temporal Utility Profiles in Demand Model

As described in detail in Chapter 3, using temporal utility profiles in the demand model is the most theoretically advanced approach, and its operationalization on the demand side requires that temporal utility profiles be defined for each activity. The attractive part of this approach is that these profiles are indeed implicitly defined in the TOD choice model embedded in any ABM. However, conversion of the time-of-choice model output into utility profiles with the necessary level of temporal resolution is not a trivial procedure and has yet to be developed and explored. The crux of the problem is that a TOD choice model produces probabilities for each activity to be undertaken at a certain time in the form of a joint start (arrival) and end (departure) time probability over all feasible combinations $P(t_a, t_d)$ such as:

$$\sum_{t_a=0}^{T}\sum_{t_d=t_a}^{T} P(t_a, t_d) = 1 \tag{5.7}$$

These probabilities are defined for each activity, and they are not directly comparable across different activities. To convert the TOD choice probabilities into temporal utility profiles, an overall scale U_k for each activity k has to be defined. The utility profile could then be calculated as

$$u_k(t_a, t_d) = U_k \times P(t_a, t_d) \tag{5.8}$$

The overall scale reflects the importance of a unit duration of each activity versus generalized travel cost. General travel cost C_{ad} is a part of the TOD choice utility $V_k(t_a, t_d)$ used to calculate the probability $P(t_a, t_d)$. Hence, the following estimate of U_k can be suggested that is essentially the coefficient of travel cost in the TOD choice utility (it is assumed that this is a single coefficient not differentiated by departure or arrival time):

$$U_k = \frac{\partial V_k(t_a, t_d)}{\partial C_{ad}} \tag{5.9}$$

However, these techniques are yet to be explored and further research is needed to unify TOD choice and temporal utility profiles. Also, even if the temporal utility profiles are available for each activity, their incorporation in an operational travel demand model is not straightforward. In a certain sense, two approaches similar to the approaches outlined above for the schedule delay method can be adjusted to the temporal profiles framework.

The first approach would employ the network simulation model to produce travel time distributions for each trip departure time bin (30 minutes). The demand model (mode choice) would then convert these distributions to estimates of activity participation loss using temporal activity profiles. This approach has never been applied, and its details have yet to be explored. The second approach would include temporal profiles in the network simulation that would require a simultaneous choice of network routes and departure times for the entire daily schedule (or each travel tour to make this model more realistic). Theoretical constructs of this type and corresponding experiments in small networks have been reported (Kim et al. 2006; Lam and Yin 2001). However, at the current time, the second approach cannot be recommended for implementation in real-size networks.

Incorporation of Reliability in Network Simulation

This section presents a concise overview of each method of quantification of travel time reliability from the perspective of its inclusion in an operational network simulation model. This means that the reliability measure of interest has to be incorporated in the route choice and generated at the O-D level to feed into the demand model.

Perceived Highway Time in Network Simulation

This method is easy to implement without a significant restructuring of the network assignment model whether a user equilibrium static assignment or advanced DTA is applied. Essentially, the generic highway travel time variable in route choice should be replaced with segmented travel time by congestion levels with the weights recommended in Table 5.3. The highway LOS skims for the demand model have to be segmented accordingly.

However, in the same way as for a demand model, perceived travel time is not a direct measure of travel time reliability for network simulation. It can be used as a surrogate when more advanced methods are not available, but it is less appealing behaviorally, and it is not the main focus of the current research.

Mean Variance in Network Simulation

This method requires an inclusion of an additional reliability term (standard deviation, variance, or buffer time) in the route choice generalized cost along with the mean travel time and cost as shown in Equation 5.9. Further on, the correspondent O-D skims for the reliability measure have to be generated to feed to the demand model (mode choice and other choices through mode choice logsums). However, implementation of

this method on the network simulation side proved to be more complicated than its incorporation in a demand model.

Any demand model, whether four-step or ABM, inherently operates with entire-trip O-D LOS measures. Consequently, adding one more measure does not affect the model structure. However, network simulation models that are efficient in large networks operate with link-based shortest-path algorithms for route choice. This results in the necessity of constructing entire-route O-D LOS measures from link LOS measures. Although mean travel time and cost are additive by link, the reliability measures are not in a general case. This represents a significant complication that has to be resolved.

Even if an explicit route enumeration is applied, which means that several entire O-D routes are explicitly considered in route choice, it is not trivial to incorporate a reliability measure like standard deviation, variance, or buffer time. In a single-run framework, this measure has to be generated based on the traffic flow versus capacity characteristics, which requires nonstandard statistical dependences to be involved. In a multiple-run framework, this measure can be summarized from multiple simulations. However, the whole framework of multiple runs has to be defined in a consistent way across demand, network supply, and equilibration parameters.

The SHRP 2 L04 project is specifically devoted to an analysis of these issues and a substantiation of the recommended methods.

Schedule Delay Cost in Network Simulation

The previous section outlines two possible approaches that differ in how and where the schedule delay cost component is calculated (see Figure 5.8).

With the first approach, schedule delay cost is calculated in the demand model as part of the mode utility calculation for highway modes. The network simulation model assigns trips based on PDT without considering PAT. The role of the network simulation model is to produce travel time distributions for each trip (through a single equilibrium run or multiple runs). Subsequently, schedule delay cost is integrated over the travel time distribution in the demand model. The most realistic implementation approach with this scheme is a multiple-run framework, which is discussed below.

In the second approach, the schedule delay cost calculation is incorporated in the network model and is fed to the demand model. Perhaps the most behaviorally appealing implementation of this approach occurs when the network simulation model is allowed to optimize departure time based on PAT and specified schedule delay penalties. This type of model can be implemented in a single-run framework; some testing of this approach has been reported (Zhou et al. 2008).

In both cases, the main (technical) obstacle for practical implementation of the schedule delay approach is the necessity to generate PAT (externally or endogenously in the demand model scheduling procedure) for each trip against which the schedule delay cost is calculated as a consequence of unreliable travel time. Further research is needed to operationalize this approach in the framework of a regional travel model.

Temporal Utility Profiles in Network Simulation

Two approaches similar to the approaches outlined above for the schedule delay method can be adjusted within a temporal profiles framework.

The first approach would employ the network simulation model to produce travel time distributions for each trip departure time bin (30 minutes). The second approach would include temporal profiles in the network simulation, which would require a simultaneous choice of network routes and departure times for the entire daily schedule (or each travel tour to make this model more realistic). Theoretical constructs of this type and corresponding experiments in small networks have been reported (Kim et al. 2006; Lam and Yin 2001).

Currently, this method cannot be recommended for implementation in real-size networks because of the many technical details that have to be explored on both demand and network supply size. However, it represents an important avenue for future research.

CHAPTER 6

Conclusions and Recommendations for Future Research

Impacts of Congestion and Pricing on Travel Demand: Behavioral Insights and Policy Implications

Variations in Value of Time Across Highway Users

Key Finding

Value of time (VOT) varies widely across the traveling population, from $5/hour through $50/hour across income groups, vehicle occupancies, and travel purposes. In addition to variation that can be explained by person and trip characteristics, there is significant situational variation (unobserved heterogeneity), with some people willing to pay almost nothing at all to save time and some (the tail of the distribution) willing to pay more than $100/hour.

Implications for Policy

The wide distribution of willingness to pay confirms that pricing can effectively serve the important function of market discrimination and demand management. Because the majority of travelers tend to have a relatively low willingness to pay, any price that affects all travelers, such as a general toll for all lanes of a highway, may influence demand at fairly modest levels. In contrast, prices of optional facilities such as high-occupancy toll (HOT) lanes and express lanes can be set at fairly high levels and adjusted to attract the desired small percentage of travelers with the highest willingness to pay. It should also be noted that in terms of revenue generation, most toll facilities in the United States are probably underpriced, and more radical pricing could be applied. However, pricing policies should be applied after a careful analysis of possible negative implications for low-income users, which is largely a function of the extent of alternative available options (transit and nontoll roads or general-use free lanes).

Implications for Modeling

In practice, most models used for travel demand forecasting have assumed a single VOT across the population for each travel purpose. In a few demand models, different cost coefficients have been used for different income groups and vehicle occupancy levels. Differentiation of VOT is even less typical in network simulation procedures, in which travel purposes and income groups are frequently lumped together. It should be noted that these (unfortunately, prevailing) practices result in significant aggregation biases that affect the accuracy of traffic and revenue forecasts. Whenever possible, the analyst should use random coefficients to estimate the distribution of VOT across the population, as depicted above. Such methods have been used in the context of forecasts for the introduction of particular tolled facilities. For more general use, newer activity-based forecasting models that use a microsimulation approach can simulate a different VOT for each person and trip, which provides the most disaggregate treatment of VOT and thus avoids one important source of possible errors and biases in the forecasts.

Related Technical Detail

The finding that there is a wide range of variation in willingness to pay for travel time savings across the population will come as no surprise to modelers or decision makers. Major studies in Europe focused on measuring determinants of VOT have consistently found significant differences related to income, mode, vehicle occupancy, travel purpose, congestion levels, and other factors. The C04 team's research confirmed the variation at an even wider range of VOT values. It also has shown that the stereotype no longer holds that the majority of highway users will be willing to pay somewhere between $10 and $15/hour of travel time saved and that this value does not change much over years. In fact, at least 50% of commuters in most metropolitan regions in the United

States will be willing to pay $20 or more per hour of travel time saved.

Furthermore, it has long been understood that there is a wide variation in VOT from individual to individual that cannot be readily related to data on person and household characteristics or trip context. Such differences can be related to individuals' personalities and to situational factors that are not available in the data. With the advent of more powerful model estimation techniques, such as mixed logit estimation, it is now possible to estimate the distribution of a coefficient in terms of the mean, variance, and shape of the distribution, rather than estimating a single-point value. Such models estimated in this study and elsewhere indicate that there is a great deal of residual variation in VOT beyond the considerable variation already explained by differences in income and other explicit variables. The best model fit is typically obtained using a lognormal distribution, like that shown in Figure 6.1, with most travelers having fairly low values (the median value below the mean value), and with relatively few travelers having quite high values (represented by the long tail). These distributions can be further segmented, with a systematically different average VOT for each segment. However, the general shape of the distribution still holds within each segment.

Income and Willingness to Pay

Key Finding

Household income and personal income have a very strong relationship with VOT and willingness to pay, but the relationship appears to be less then linear. To account for the income effect, cost variables in travel models (including tolls) should be divided by household income, raised to a power in the range 0.6 to 0.8 depending on the trip purpose. As an example, when using a power of 0.7, if income is doubled, VOT increases by 62%; if income is halved, VOT decreases by 38%.

Implications for Policy

Although income certainly is not the only factor that influences willingness to pay for travel time savings, the income effect is quite strong, so many of the benefits of pricing will tend to

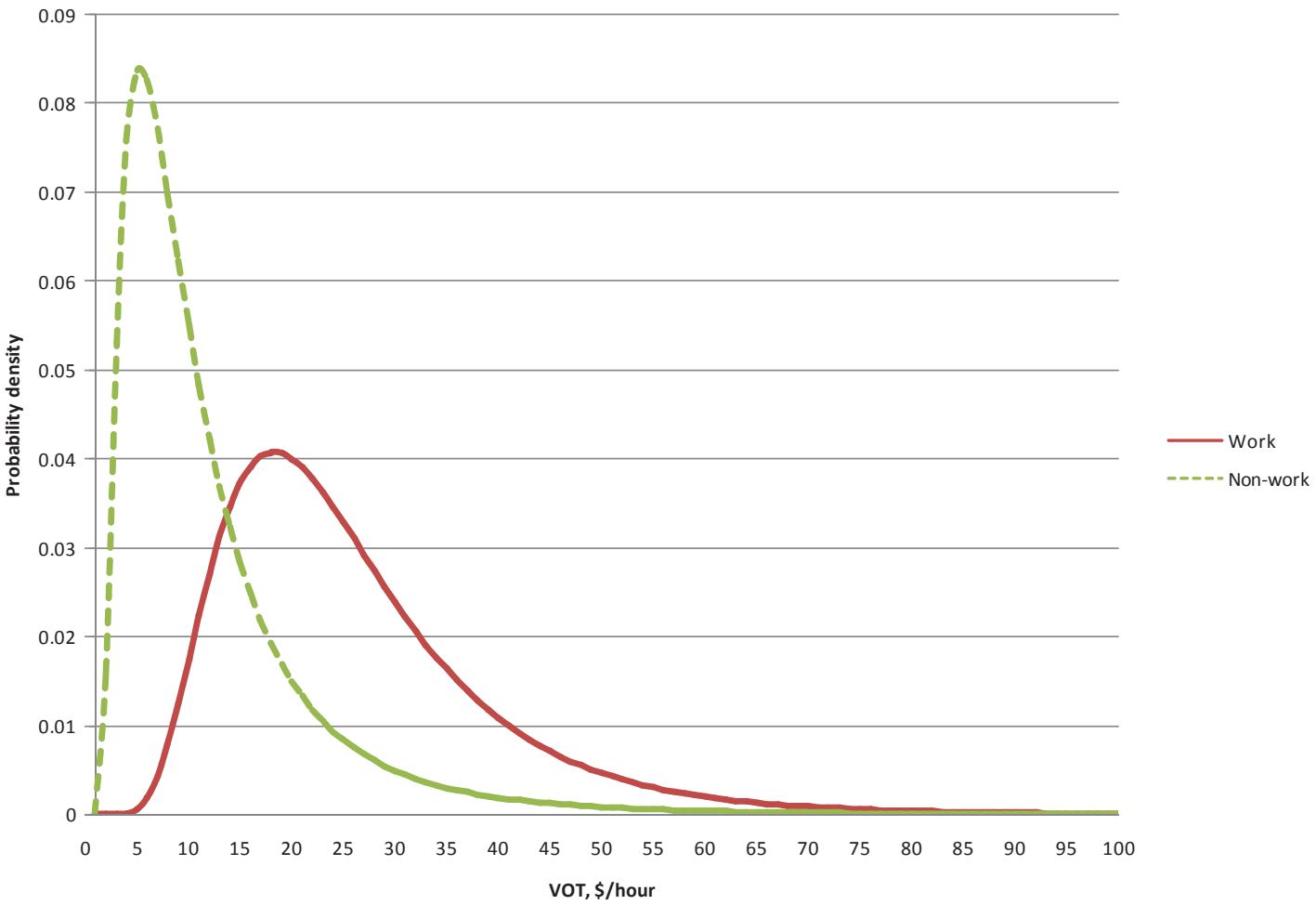

Figure 6.1. Typical lognormal distribution for VOT in the United States.

be purchased by those who can most afford them, and equity considerations cannot be discounted. Of course, lower-income individuals can also derive benefits in the form of increased options, as well as improvements in traffic conditions if capacity in the entire system can be increased through priced facilities. An important factor that mitigates income effect is the parallel effect of car occupancy. As discussed below, low-income commuters have more opportunities than high-income commuters to carpool and share commuting costs. In this sense, not only transit, but also high-occupancy vehicle (HOV) and HOT lanes represent viable alternatives for low-income travelers.

If estimates of VOT are to be used in social cost–benefit analysis, there is debate among economists about whether it is appropriate to value benefits differently for different income groups. This is a normative issue that is outside the scope of this project. However, if the benefit component includes user benefits (as most economists recognize), then the entire highway utility function can be used as the basis for calculation with income and other effects.

Implications for Modeling

In recent practice, many forecasting models have not included income as a moderating influence on travel cost sensitivity. When income is considered, it is typically either used in a simplified linear form to scale travel costs or as a segmentation variable, with different cost coefficients in different income ranges (or different "bias" constants). Although those two approaches often approximate the one recommended here, neither approach seems entirely appropriate. The assumption of linearity with income seems too strong, particularly in higher income ranges, and the piecewise linear approach often results in strong nonlinearities or discontinuities (or both) in the effect of income and does not have a strong statistical or behavioral basis. The recommended approach is empirically justified across a wide body of evidence and provides a smooth response surface for forecasting.

Related Technical Detail

The graph in Figure 6.2 shows the relationship between income level and VOT that arises from various estimates of the income exponent *e* in the utility formulation. The best fit at an exponent in the range 0.6 to 0.8 indicates that the relationship is less than linear with income, but still very substantial. Furthermore, the sensitivity of VOT to income is greater in lower income ranges than it is in higher income ranges. One possible reason for this is that at lower income levels, budget constraints may be quite strong, and certain travel options may be simply unaffordable. As income increases, however, the likelihood that any travel option is truly unaffordable becomes less, and the budget effect becomes less a matter of constraints and more a matter of preferences between different types of expenditures, as well as time spent at home versus time spent out of home for discretionary activities. In terms of travel behavior in general and willingness to pay for travel

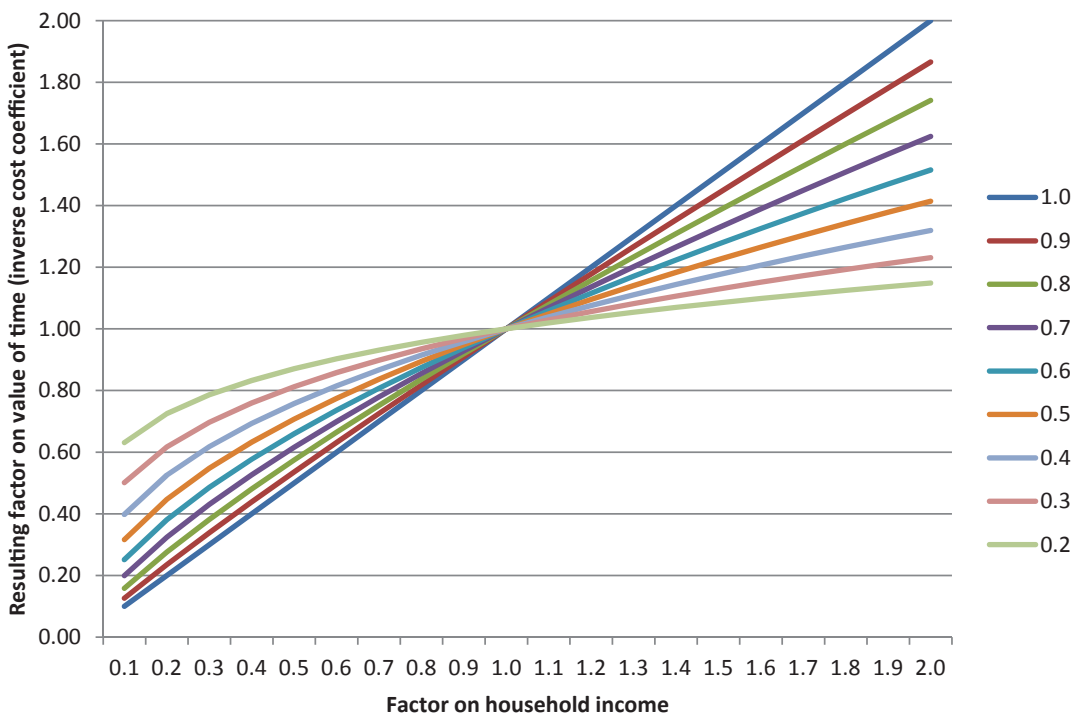

Figure 6.2. Effect of income exponent on VOT.

time savings in particular, there is not much of an income impact between, say, a household with a $200,000 income versus a household with a $400,000 income that is expressed in a lower income exponent *e* that makes the curve flatter.

This variable is useful for examining expected behaviors when future conditions change. For example, if transportation costs rise drastically with the cost of fossil fuels or new forms of pricing, then the price of travel relative to income would shift, so that even higher-income households could face the sort of budget constraints that are now common for lower-income households. In that case, one might expect a more linear relationship of willingness to pay with income. In modeling terms, that suggests assuming an exponent on income closer to 1.0.

Auto Occupancy or Group Travel and Willingness to Pay

Key Finding

Auto occupancy has a very strong estimated relationship with VOT and willingness to pay, reflecting in part cost sharing between the driver and passengers. The relationship appears to be slightly less than linear. To account for occupancy effects, cost and toll variables in travel models should be divided by occupancy raised to a power in the range of 0.7 to 0.8. As an example, using a power of 0.8, if occupancy increases from one to two, VOT is multiplied by a factor of 1.74; if occupancy is increased from one to three, VOT is multiplied by a factor of 2.41.

Implications for Policy

The fact that a group of vehicle occupants is, on average, willing to pay more than a solo driver in the same choice context suggests that a tolled facility should attract a higher percentage of multioccupant vehicles than a free facility will, even if no special discount is offered for carpools. Looked at in another way, a carpool discount is being offered to a group that tends to value it the least. In purely behavioral terms, this situation is similar to offering a discount to higher-income drivers. On the other hand, ridesharing is advantageous in terms of increasing system capacity, and the conversion of HOV lanes to HOT lanes may potentially reduce or discourage carpooling among individuals with higher VOT by offering solo drivers the same travel time advantage without the inconvenience of ridesharing. From that standpoint, offering free or discounted use of toll lanes to carpools will at least provide an incentive for carpoolers to continue ridesharing, even if it does not attract a great deal of additional ridesharing.

There is an important objective difference between carpooling opportunities for different income groups that has to be taken into account in policy evaluations. In general, low-income commuters have a higher probability of forming a carpool (i.e., finding a partner) for the following reasons:

- Low-income workers normally have a fixed work schedule that simplifies carpooling logistics; high-income workers are characterized by more flexible work schedules that make carpooling arrangements difficult;
- Low-income workers tend to live in dense residential clusters where collecting and distributing passengers require minimum extra time. High-income workers tend to reside in low-density suburbs where this extra time might be significant; and
- Low-income jobs tend to form clear clusters of multiple jobs, but high-income jobs might be more specifically distributed (e.g., near major universities). This factor may vary depending on the structure of the metropolitan area and its core; for example, there may be large clusters of high-income jobs in the central business district (CBD).

In general, the higher opportunity for carpooling for low-income workers mitigates the equity concerns regarding pricing because the cost can be effectively shared within the carpool. In the presence of significant tolls, high-income workers can only switch to transit, but low-income workers can switch to either transit or HOV. This consideration is frequently missing in policy analysis of pricing projects, which may result in an exaggeration of equity concerns.

Implications for Modeling

Dividing travel cost by vehicle occupancy is already a fairly standard practice in applied modeling, so no major change in practice is required in this respect. The team's main recommendation is to divide costs by a function of occupancy that is somewhat less than linear, rather than assuming strict linearity. With respect to the combined income–occupancy effects, it is important to have income-specific components in the car occupancy choice that reflect differential opportunities to carpool by income. Simplified approaches based on average occupancy coefficients tend to mask these important effects and portray pricing projects in an extreme way with respect to different income groups.

Related Technical Detail

The graph presented in Figure 6.3 indicates the effect on willingness to pay as vehicle occupancy increases from one (drive alone) to higher occupancy levels. The recommended approach for modeling is to divide travel cost not only by a power of income, as described above, but also by a power of occupancy *f* in the range of 0.7 to 0.8. These different effects can be introduced simultaneously because they arise for

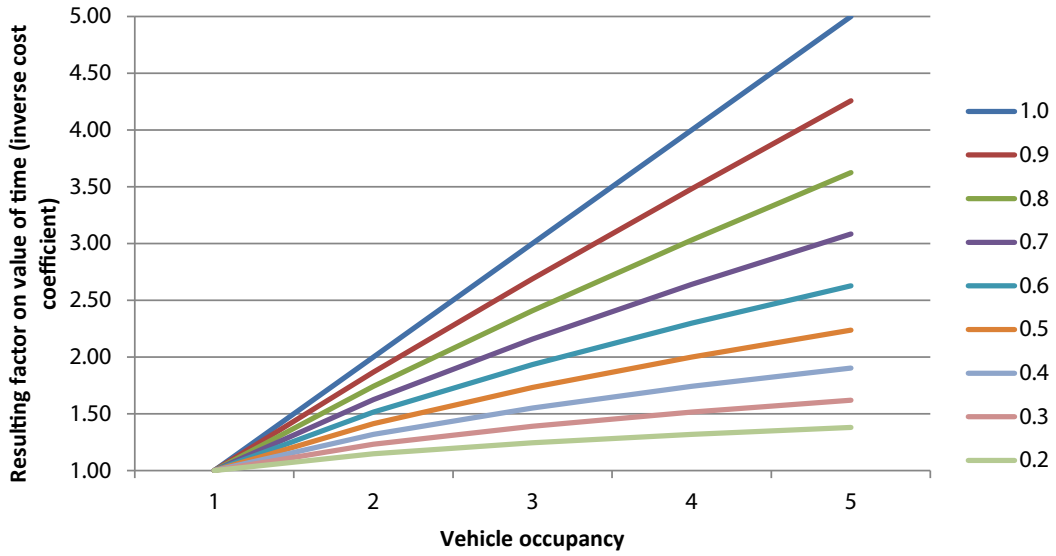

Figure 6.3. Effect of car occupancy exponent on VOT.

different reasons. Income effects are due primarily to monetary budget constraints, but occupancy effects are due primarily to the possibility of cost sharing among occupants. The estimated effect is somewhat less than linear, which may be due to the reason stated above for income: as the cost for each additional occupant becomes smaller, it is essentially a smaller fraction of each occupant's disposable income and less likely to be severely restricted by budget constraints.

Other aspects of vehicle occupancy also influence willingness to pay. For example, the monetary considerations for a commuting carpool consisting of coworkers may be different from those of a number of household members traveling together for a nonwork trip. Another consideration relates to the travel party composition. Adults would most probably share cost. However, on trips in which an adult would escort children, cost sharing is less logical. All these effects are analyzed in more detail in Chapter 4. Empirically, however, the effect of occupancy on VOT seems to be quite similar for work and nonwork purposes. This might be a manifestation of the fact that joint travel for nonwork purposes is frequently associated with fixed-schedule events (like going to a concert or theater or visiting a doctor) and other activities of relatively high priority; hence, the willingness to pay is also higher.

It is important to recognize the strength of the car occupancy effects in the context of different pricing forms. Even with a fixed toll per vehicle, carpools have a significant advantage in terms of VOT that is expressed in multipliers of 1.7 for HOV-2 and 2.4 for HOV-3. This effect can be combined with the toll differentiation by occupancy. Consider, for example, an elaborate HOT-4 lane policy in which a single-occupant vehicle has to pay a full toll, HOV-2 vehicles have to pay a half toll, HOV-3 vehicles have to pay a third of the toll, and HOV-4 vehicle can use the lane for free. Taking into account the higher willingness to pay for carpools, the equivalent toll multiplier for HOV-2 will constitute $1/(2 \times 1.7)$, which is less than a third. For HOV-3 it will constitute only $1/(3 \times 2.4)$, which is less than a seventh. This has important policy (mitigating) implications, especially for low-income commuters.

Constraints on Time-of-Day Shifting: Carpools and Single-Occupant Vehicles

Key Finding

Although commute carpools generally have a higher VOT, they also tend to have tighter scheduling constraints and tend to be less flexible in their capacity to shift departure time away from the peak period and hour.

In the departure-time choice models estimated as part of this project, the team consistently found that commuters who share rides are more likely to travel in the heart of the peak periods, relative to those who drive alone. Those in carpools need to coordinate their commute schedules with cotravelers, so it is less likely that they can adjust their departure times earlier or later to avoid peak congestion or pricing. In other words, it is easier to find partners for carpooling for conventional commuting schedules than for earlier or later schedules.

Implications for Policy

Compared with solo drivers, carpoolers on average are less able to retime their trips away from the peak congestion times.

This means that time-of-day (TOD) pricing and other peak-spreading policies will tend to be less successful in influencing the behavior of carpoolers. As mentioned above, it may be important to design policies to avoid inadvertently discouraging ridesharing by maintaining some level of travel time or price advantage (or both) for HOVs, as in case of HOV and HOT lanes. A related consideration is that congestion-pricing policies could be more effective if they were accompanied by policies encouraging employers in the CBD (or other relevant congestion pricing zone) to shift working hours in a preplanned way, as well as introduce flexible or compressed work weeks.

Implications for Modeling

Modeling studies intended to predict peak-spreading behavior and responses to TOD pricing should include different sensitivities for different car-occupancy levels. In general, the propensity to switch from the peak hour to a different hour should be inversely proportional to vehicle occupancy.

Importance of Value of Reliability and Relationship to Value of Time

Key Finding

Improvements in travel time reliability are at least as important as improvements in average travel time. The reliability ratio (the value of reducing the standard deviation of travel time by 1 minute divided by the value of reducing the average travel time by 1 minute) is estimated in the range of 0.7 to 1.5, and trends in results from other research suggest this value is increasing.

A great deal of the effort in this project was devoted to deriving estimates of the value of reliability (VOR) from real-work data on actual choices, simultaneously with estimates of VOT and other time- and cost-related effects. Although it has proven very difficult to assemble adequate origin–destination (O-D) level-of-service (LOS) data for such modeling, the team has been able to generate reliability skims and produce estimation results that make sense behaviorally and are fairly consistent with prior evidence. In general, it appears that travelers value variation in travel time reliability (day-to-day variability) at least as highly as variations in the usual, typical travel time. Although the team tested various ways of specifying the variability variable (including standard deviation in day-to-day time, the difference between the 90th and 50th percentile times, and the difference between the 80th and 50th percentile times), the measure that produced the most consistent results was the standard deviation in travel time divided by journey distance.

When the team evaluated the estimation results to impute the reliability ratio (VOT/VOR for an average trip distance), ratios in the range 0.7 to 1.5 for various model specifications were obtained. In prior published work, much of it based on stated preference (SP) studies from Europe, typical values are in that same range for auto travel, with higher ranges up to 2.5 for rail and transit travel. The SP results, however, indicate that estimates may vary a great deal depending on how the reliability concept is presented to respondents in the hypothetical scenarios. This variability in SP estimates is a reason why it is so crucial to obtain new estimates based on actual choices at the trip (O-D) level. The results of this project provide an important step in that direction.

Implications for Policy

Highway investments that improve travel time reliability will tend to be just as beneficial for travelers as investments to reduce typical travel times. This finding underlines the importance of addressing key bottleneck points, of using transportation systems management and intelligent transportation systems to monitor and adapt to congestion levels on the network, and of using systems that avoid nonrecurrent congestion and recover from such congestion as quickly as possible. For managed lanes and other priced facilities in particular, the "guarantee" of a reliable travel time may be of great value. This makes variable pricing, and especially dynamically priced lanes, one of the more effective pricing forms that is at the same time very attractive to the user. This also emphasizes the importance of effective accident management, because the consequences of traffic accidents constitute a significant share of long delays.

Implications for Modeling

Although the team has shown that it is possible to estimate models using measures of day-to-day travel time variability from real and simulated highway networks, further progress will need to be made before this method is feasible for most travel demand forecasts, particularly in terms of widespread collection of data for actual levels of travel time variability at the O-D level. Certain technical issues in network simulation must also be resolved, specifically the incorporation of travel time reliability in route choice and the generation of O-D travel time distributions (reliability skims) instead of average travel times.

Other SHRP 2 projects, such as L04 (Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools) and C10 (Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network), are aimed at bringing these methods into widespread application. In the near term, it may be most applicable for corridor-level and facility-level forecasts. Some simplified implicit measure of reliability (such

as perceived highway time by congestion levels, as explained below) can also be applied with the existing model structures and network simulation procedures.

Effect of Travel Distance on Value of Time and Value of Reliability

Key Finding

Savings on average or typical travel time (VOT) are valued more highly for longer trips than for short trips, except for a special effect on very long commuting trips (over 40 miles). For VOR, there is a relative damping effect for longer trips. These findings suggest the efficacy of using higher-priced managed lanes to address key bottlenecks in combination with lower distance-based tolls on the wider highway network.

Implications for Policy

Traffic bottlenecks tend to add a great deal of variability (unreliability) to all trips that pass through them, regardless of the total trip distance, and the results indicate that all travelers affected by such chronic variability will derive considerable benefit from making the system more reliable. In contrast, improvements that increase average speeds or reduce travel distances without substantially improving reliability will not be valued very highly by those who only use the facility for a short distance. The implication is that distance-based tolls are appropriate in general, but higher prices that are not based on distance may be more appropriate to address key bottlenecks.

Implications for Modeling

The analysis results indicate that both VOT and VOR tend to vary with O-D trip distance. Using a constant VOT and VOR for a wide range of short and long trips is yet another unreasonable simplification pertinent to most travel models. For the most accurate predictions, distinctions in VOT and VOR values should be used in demand forecasting models.

Related Technical Detail

The impact of trip length on VOT (and possibly VOR) has been analyzed in several interesting studies from both theoretical and empirical perspectives. There is no full consensus regarding the direction of impact, and in most models used in practice, VOT is considered fixed (and VOR is ignored). Positive, negative, and nonmonotonic effects all have been considered and found at least with some data sets and forms of analysis. However, probably in the majority of previous studies, the authors arrived at the conclusion that VOT should grow monotonically with trip length for the following reasons:

- Cost Damping. This effect can be generalized as a diminishing marginal disutility of travel cost with the growing distance. One plausible explanation is that travelers have a relative rather than absolute perception of the cost of a trip. That is, ±$1 is perceived as a small difference if the base cost is $10, but it is a crucial difference when the base cost is only $2. Another realistic explanation is a poor perception of car-operating cost versus (out-of-pocket) parking cost and tolls. While car-operating cost is proportional to distance, parking cost and tolls usually are not (unless mileage-based pricing is applied). Yet another reason might be cheaper housing and higher disposable income for long-distance commuters. Additionally, for nonwork travel, trip frequency in many cases is inversely proportional to trip length. That is, longer-distance travel is rarely associated with a special event (like major shopping for furniture versus routine grocery shopping), when willingness to pay might be higher. Also, in models in which car occupancy is not accounted for explicitly, higher car occupancy (and the corresponding higher willingness to pay) can be correlated with longer trips; and
- Time Valuing. This effect can be generalized as a growing marginal disutility of travel time with growing distance. The most basic explanation for this is a rigid time budget constraint that almost every person has: 24 hours each day, most of which goes to basic sustenance and mandatory activities like work and school. As the result, the longer the travel, the more valuable each minute of the travel time savings becomes. Other explanations suggested in previous research include risk aversion if reliability is not accounted for explicitly (longer trips might have a higher uncertainty in terms of arrival time) and unfamiliarity with distant locations (and associated route, location, and parking searches).

All these effects are analyzed in detail in Chapter 4. In reality, these multiple reasons work altogether, and in modeling terms they are captured by the distance-based multiplier on travel time (described above), which also proportionately affects VOT. The entire additional multiplier on the travel time that collects all distance terms is of special interest because it directly expresses the impact of distance on VOT. The results are shown for work-related purposes in Figure 6.4.

The shape of the distance-effect curves is similar to the shape reported in previous research, although in most sources, only monotonic functions were obtained. Depending on the market segment and other model components, the inverted U effect can be less or more prominent, with a very small impact on the overall model fit.

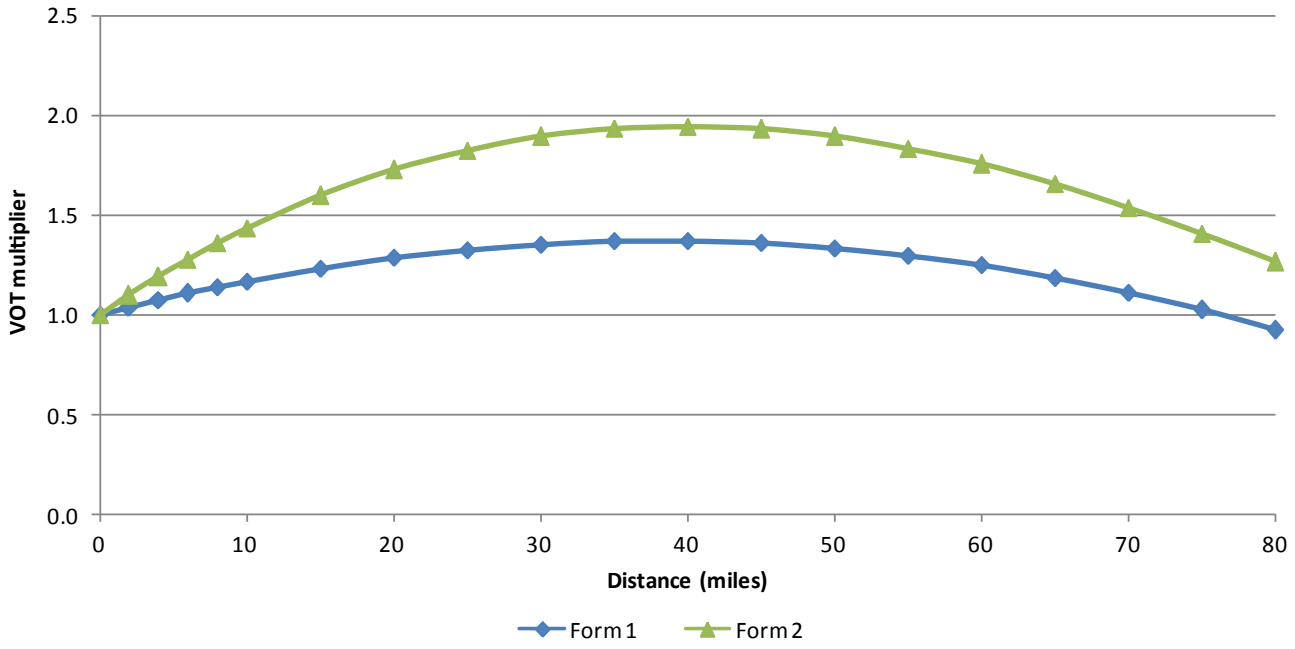

Figure 6.4. Travel distance effect on VOT: work-related trips.

The team believes that the lower VOT for long-distance commuters is a manifestation of restructuring the daily activity–travel pattern. In particular, long-distance commuters tend to simplify their patterns and not have many additional out-of-home activities on the day of their regular commute because the work activity and commuting consume most of the daily schedule. To compensate for this, long-distance commuters tend to have compressed work weeks or telecommute more frequently, which gives them an opportunity to combine nonwork activities on one particular day of the week (most frequently, Friday) when they do not commute to work. In contrast, short-distance commuters tend to have multiple additional out-of-home activities that add pressure to the daily schedule. In a certain sense, there are also lifestyle and residential self-choices embedded here. That is, long-distance commuters are willing to sacrifice out-of-home nonwork activities for better living conditions (and presumably more intensive in-home activities).

An additional factor that may result in a higher tolerance to long travel times for commuters is the possibility of using the commuting time productively (especially if convenient transit modes like commuter rail are used). Using cell phones and laptops or reading a newspaper or book reduces the burden of travel time. This is somewhat less relevant for auto trips, although cell phone usage in auto travel is becoming quite common, as well.

For reasons discussed above, travelers seem to value each minute of typical, average travel time on longer trips, when the total amount of time saved can substantially reduce the time needed for the trip. For reliability, however, the team obtained the opposite result. Differences in travel time variability appear to compensate over longer journeys, and travelers place more value on each minute of variability for shorter trips. This is reflected in the reliability term in a form of standard deviation of travel time scaled by distance that makes VOR inversely proportional to distance.

Evidence of Negative Toll Bias

Key Finding

There is a significant negative threshold bias against paying a toll, regardless of the toll amount. This preference against paying a toll is generally supported across travel purposes, as found in both revealed preference (RP) and SP data, and is also supported by research in behavioral economics. The estimated toll penalty effect for auto trips is generally equivalent to as much as 15–20 minutes of travel time.

Implications for Policy

The resistance to paying a toll appears to present an obstacle to the effective widespread introduction of congestion-pricing policies. In many cases, however, a pricing policy can be effective even if only a limited proportion of drivers choose to pay the toll, and just like VOT, the resistance to paying any toll at all may vary a great deal across the population. In that sense, toll bias becomes another dimension of market discrimination, similar to VOT. What is important is that resistance can be overcome by a guaranteed superior level

of service in terms of travel time savings and improvements in reliability. In this sense, tolling existing facilities in order to collect revenue, but without a substantial LOS improvement, would generally be perceived very negatively by highway users.

Furthermore, one can expect that resistance to paying a toll will fade over time as road pricing becomes more ubiquitous and more convenient. In the past, drivers had to wait in lines to pay, which in itself could explain a good deal of resistance to tolls. Now, with the introduction of electronic tolling, which has become commonplace in many regions, paying the toll is both faster and less noticeable in terms of the amount of money actually being spent. There are already fully automatic open-road toll collection technologies in place that completely eliminate toll delays. The more widespread that electronic road pricing becomes, the more policy makers can expect antitoll bias to be reduced, although it may never disappear completely for some travel segments.

Implications for Modeling

In practice, there are different opinions and methods regarding the incorporation of antitoll threshold terms in forecasting. Sometimes they are avoided in forecasting on the basis that they are not rational in economic terms. Empirically, however, antitoll threshold terms do appear to be real, so they should be included to obtain the most accurate results, at least for short-term forecasts. In general, this bias would result in a more conservative traffic and revenue forecast if travel time savings are insignificant, but it also may result in a more optimistic forecast for pricing projects that improve travel time significantly. For longer-term forecasts, it may be appropriate to explore scenarios with reduced or eliminated antitoll bias threshold terms.

Related Technical Detail

After accounting for differences in price, average travel time, and reliability, there appears to be a general reluctance in the population to paying any toll (a toll bias) to use a highway facility. This result is frequently obtained in SP studies, in which it is sometimes explained as a "protest response" or "strategic bias" to avoid the introduction of tolls. However, such a bias is also found in RP data, as it is in models estimated for this project. In particular, a toll bias was confirmed by the RP data from New York, where toll facilities have a long history and explanations like short-term psychological protest or ramp-up cannot be applied.

Although the relative size of such a toll bias tends to be smaller when estimated from RP data as compared with SP data, it can still be substantial, and equivalent to as much as 15–20 minutes of travel time. In other words, travelers would go that far out of their way to avoid paying any toll at all. This type of behavior has also been noted in recent texts in behavioral economics, which note that people are observed to go to seemingly irrational lengths to get something for free as opposed to paying for it (Ariely 2010).

It is actually logical to have a significant toll bias in combination with a relatively high willingness to pay as measured by VOT. These two factors are screened separately in the highway utility of the form adopted in the current research. In a simplified form in which the toll bias is not included, the entire utility gets readjusted, which most frequently results in a lower VOT. This type of result is illustrated in Figure 6.5.

In Figure 6.5 it is assumed that the toll value is fixed, and the relative utility of toll option versus nontoll option (both options are assumed available for the user) is analyzed as a function of travel time savings achieved with the toll option. If there is no time savings, the relative utility of the toll option is logically negative. For a model without a toll bias, the associated disutility is equal to the toll value in equivalent units of utility. For a model with a toll bias, the associated disutility is even worse because it includes both the toll equivalent and bias.

The point at which the difference between toll and nontoll utilities becomes zero corresponds to the 50–50 split between toll and nontoll users. For the model without a toll bias, this point corresponds to the time savings equal to the toll value divided by VOT. For the model with toll bias, this point is shifted and corresponds to the toll value divided by VOT plus toll bias equivalent in minutes. By virtue of the model estimation on the same data set, the model with a toll bias would have a greater slope (and higher VOT).

As Figure 6.5 shows, the response of a model with a bias and higher VOT to pricing policies can be very different from the response of a simplified model without the bias and adjusted (lower) VOT. A model with bias would tend to produce a conservative traffic and revenue forecast until substantial time savings are guaranteed for the toll users. However, when the savings grow, the number of toll users will grow at a higher rate. In contrast, a simplified model would overpredict the number of toll users if the travel time savings are insignificant, but would underpredict the number of toll users when the travel time savings grew significantly. In a certain sense, the model suggested in the current research would be more demanding from the pricing projects to guarantee a value for money.

Hierarchy of Likely Responses to Changes in Tolls and Congestion

Key Finding

Traveler responses to congestion and pricing depend on the range and attractiveness of available alternatives. From the

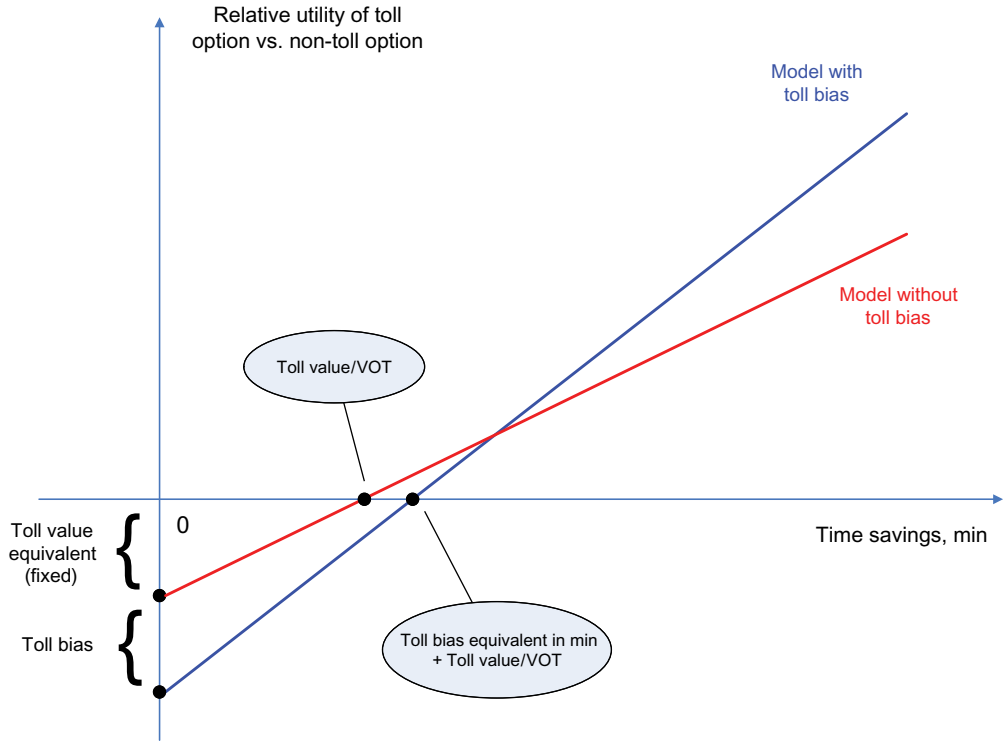

Figure 6.5. Effect of negative toll bias.

highest to the lowest propensity to change behavior, these responses are as follows:

- Primary. Change lane or route type or make minor shifts in departure time (up to 1 hour earlier or later), or both;
- Secondary. Switch between auto and transit (in transit-rich areas) or change car occupancy (carpooling), or both;
- Tertiary. Cancel, relocate, or reschedule most flexible and discretionary trips and activities (or some combination of these changes); and
- Longer Term. Change the location of home, work, or other important activity; change the number or types of vehicles owned.

The models estimated for this project covered a range of travel choices. When possible, nested hierarchical models were estimated to determine which types of choices are most sensitive to travel time and cost changes. The highest propensity to change appears to be between tolled and nontolled lanes or routes. A change of route requires little effort and little or no adjustment in travel schedules, and the choice can even be made en route subject to perceived traffic conditions at a specific point in time. Travelers also show a fairly high propensity to make minor shifts in departure time of an hour or less, since the smaller the shift, the less rescheduling of activities that is required, and the more familiar the traveler is likely to be with the typical traffic conditions over time.

Somewhat less likely are changes in either travel mode or car occupancy. These may include switching between auto and transit in areas where transit services are competitive and may also include switching between driving alone and ridesharing when cotravelers can be found. Mode shifting is most prevalent for commute trips and other very frequent trips for which information about transit services or possible carpoolers is most available or worth investigating. Monthly transit pass in regions like New York offers significant savings compared with a single-ride ticket, which also makes switching to transit most logical for daily commuters.

Less likely responses to changes in congestion or pricing are changes in the choice of destination locations, the rescheduling of trips to very different times of day, or changes in the frequency of making trips from home. These types of changes are the least likely for activities that are most constrained in time and space (e.g., work and school trips or medical appointments). For more flexible and discretionary types of trips, these types of shifts may actually be more likely than changing the mode of travel.

Finally, in the longer term, people may make more substantial changes as opportunities arise and life-cycle transitions occur. These shifts include changing the number or type (or both) of vehicles owned and the location of home, work, school and other key travel anchor points relative to one another. Although the team did not model such choices as part of this study, other research has indicated that the speed

and cost of traveling by car can have a marked influence on such decisions, even if they are not the primary decision factors. In Chapter 3, the team outlines an approach to modeling a wide range of possible longer-term responses to congestion and pricing by means of accessibility measures that are derived from the estimated primary choice of route, mode, and TOD.

Implications for Policy

Decisions influencing traffic congestion and the cost of driving can affect travel behavior in a number of different ways, and the relationships are often complex and can shift over time. This aspect of travel behavior argues for using advanced demand-simulation models to guide policy, rather than relying solely on mental models and experiences. The most predictable effects tend to be those that require only minor adjustments on the part of travelers, such as choosing a new tolled facility adjacent to an existing facility or choosing to travel at a slightly different TOD. In terms of making a pricing policy more effective in tackling congestion, an important factor relates to the presence of competitive alternative modes and destinations. In this sense the worst situation occurs when the job clusters and main nonwork attractions are concentrated in the CBD area, but the transit service is very limited. Unfortunately, this situation is typical of many metropolitan regions in the United States, and in this case, even a radical pricing policy would hardly be expected to resolve the congestion problem, and it could instead generate wide public anger. Pricing policies are most effective in combination with transit improvement and smart land use development.

Implications for Modeling

Modeling systems should be able to represent the influences of travel time and cost on all of the types of decisions listed above, and the models should be integrated so that appropriate relative sensitivities are reflected at the different hierarchical levels. These relative sensitivities should also allow for variation in travel segments and different types of individuals. In practice, this will require an activity-based microsimulation model, ideally used in combination with accurate dynamic simulation of traffic congestion, such as the one being developed in SHRP 2 C10A, Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network.

Summary of User Segmentation Factors

Key Finding

Many factors can affect VOT, VOR or traveler responses to congestion and pricing, such as person, household, land use, and travel characteristics. These responses are also subject to many situational constraints. It will never be possible in regional travel models designed for long-term forecasting to account for all the details of user characteristics implicit in traveler response. It seems right, however, to account explicitly for the most important and systematic effects, and also to apply reasonable assumptions about the probabilistic distributions of VOT and VOR in order to account for the residual heterogeneity.

Implications for Policy

Most of the important effects that affect traveler responses to congestion and pricing are highly differentiated by highway user groups. When the user benefits are calculated and winners and losers are identified, the analysis has to be implemented with a necessary user segmentation that at a minimum should include trip purpose (work and nonwork), income group (three to four categories), car occupancy (three to four categories), commuting distance (two to three categories), and household size (two to three categories). In addition, it is highly desirable to account for significant unobserved user heterogeneity and situational variability by applying probabilistic rather than deterministic VOT/VOR. Simplified methods that operate with a crude average VOT/VOR are subject to significant aggregation biases and will generally not portray a pricing project in an adequate way.

Implications for Modeling

For accurate policy evaluation, modeling systems should be segmented according to the main effects described above. In this regard, traditional four-step demand models and static traffic assignments, which are still the most common tools in practice, hold very little promise, because limited segmentation is one of the major constraints of these models. Also, it is practically impossible to incorporate distributed parameters in these aggregate constructs. Activity-based models (ABMs) on the demand side and dynamic traffic assignment (DTA) on the network simulation side offer the potential for significantly better platforms for modeling highway congestion and pricing because they are based on the concept of individual microsimulation.

Related Technical Detail

Table 6.1 summarizes a wide range of segmentation dimensions explored in the current study. Some of major factors like travel purpose, income, and car ownership are discussed above. For other factors, the team reports their experience and recommends how these factors can be incorporated in travel models in the near future, as well as the potential for future research. In Table 6.1, each possible dimension

Table 6.1. Highway User Segmentation

Dimension for User Segmentation	Previous Research	Current Study	Future Research
Socioeconomic Segments of Population by			
Household income	Positively correlated with VOT (frequently linearly)	Positively correlated with VOT (weaker than linearly but with a constant elasticity of 0.6–0.8)	Elaborate income variable (disposable instead of gross); incorporate budget constraints explicitly
Person age	Higher VOT for middle age (sometimes for females only)	Not significant statistically for VOT; younger adults have higher preference for transit and nonmotorized modes	Elaborate age effects in walk and transit access variables
Gender	Females have a higher VOT because of busier daily schedules	Females have somewhat higher VOT especially in presence of a preschool child	Link gender effects to household composition and roles
Worker status	Workers have higher VOT for nonwork travel than nonworkers because of busier schedules	Could not separate worker status effect from trip purpose effect for VOT/VOR; workers have higher preferences for solo driving in mode choice compared with nonworkers; full-time versus part-time affects TOD choice	Analyze entire-day (or multiday) patterns with respect to VOT and VOR
Student status	University students have lower VOT and higher propensity to use transit	Could not separate student status effect from trip purpose effect for VOT/VOR	Analyze entire-day (or multiday) patterns with respect to VOT and VOR
Household size	Large households are more likely to carpool	Large households are more likely to carpool	Explicitly model carpooling mechanisms
Car ownership or relative car sufficiency versus number of drivers	No direct impact on VOT/VOR except for transit captives; strong impact on mode availability and preferences	No direct impact on VOT/VOR; strong impact on mode availability and preferences	Better integrate highway route choice (auto users only) and mode choice (all travelers including transit captives)
Presence of children	Impact on VOT/VOR inconclusive	Females have somewhat higher VOT in presence of a preschool child; significant impact on TOD choice for workers	Explicitly model carpooling mechanisms and escorting
Travel and Activity Segments by			
Travel purpose and activity type	Work trips have a higher VOT/VOT than nonwork trips; special types of trips with high VOT/VOR include business, to airport, medical appointment, sporting and other fixed schedule events	Work trips have a higher VOT/VOT than nonwork trips; more detailed analysis by trip purpose was inconclusive	Analyze underlying mechanisms of behavior and activity characteristics such as schedule flexibility and situational time pressure
Weekday versus weekend	VOT/VOR is systematically lower on weekends	Analysis was limited to weekdays	Analyze weekday versus weekend with situation variables and time pressure to determine the reason for differences
Trip frequency	VOT can be higher for infrequent trips associated with special events	No conclusive results	Analyze multiday activity patterns
TOD	a.m. has highest VOT/VOR, followed by p.m. Off-peak has lowest VOT/VOR	No significant difference between TOD periods if travel time reliability is accounted for explicitly	Explicitly model individual daily schedule with schedule flexibility constraints and time pressure
Vehicle occupancy and travel party composition	VOT proportional to vehicle occupancy	VOT proportional to vehicle occupancy but weaker than linearly (constant elasticity of 07–0.8)	Analyze effects of carpool type (intrahousehold versus interhousehold) and travel party composition (adults versus adults with children)

(continued on next page)

Table 6.1. Highway User Segmentation (continued)

Dimension for User Segmentation	Previous Research	Current Study	Future Research
Trip length or tour distance	VOT grows with distance (although weaker than linearly) because of marginal cost damping and time valuing	For commuting to work, VOT grows with distance but drops for distances over 40 miles following an inverse U shape; for nonwork trips no significant effect; distance-based biases are significant for rail modes in mode choice	Explicitly account for time and budget constraints
Toll payment method	Electronic payment is favored by users beyond direct time and cost consideration because of a different perception (not out-of-pocket)	Was not possible to explore with the available data sets	Can be explored with new data sets but is probably not worth pursuing because of the wide adoption of electronic payments in near future
Situational context; time pressure versus flexible time	Limited RP evidence and discussion on VOT/VOR; significant differences in VOR when measured as schedule delay penalty in SP studies; very high VOT for trips to airports	Time pressure measures were significant in TOD choice; no conclusive results on VOT/VOR	Explore VOT/VOR in the context of entire individual daily pattern; more detailed segmentation of trips and activities by schedule flexibility
Highway travel time segmentation by facility type	Time spent on highways and freeways is less onerous than on arterial and local roads	No statistically significant results with using static assignment skims	Time coefficient differentiation by facility type should be revisited with actual data on O-D trajectories
Highway travel time segmentation by congestion level	Time savings under congestion conditions are valued 1.5–2.5 more than savings of free-flow travel time	Significant weights (1.5–2.0) on congestion delays versus free-flow travel time if reliability is not accounted explicitly	Explore time weights by congestion levels to serve as proxy for reliability; useful for simple models in practice

for user segmentation (rows) is described in terms of three aspects (columns): (1) reported results from previous studies, (2) findings from the current study, and (3) suggestions for future research. The following main groups of dimensions are distinguished:

- Socioeconomic Segments of Population. These characteristics are exogenous to all activity and travel choices that are modeled in the system. Thus, the corresponding dimensions can always be applied for any model, either for a full segmentation or as an explanatory variable in the utility function; and
- Segmentation of Activities and Travel. These characteristics are endogenous to the system of travel choices. In the mode estimation they have to be carefully related to the model structure to ensure that all dimensions or variables used in each particular model have already been modeled in the model chain.

In many respects, the current research confirmed effects reported in previous studies. However, there were several particular aspects, like the effect of trip distance or car occupancy on VOT, for which new results were obtained that differed from the previous studies. Although there is a wealth of published studies on VOT, there are only a few recent studies on VOR. Most of them are based on single-trip SP experiments, and VOR is rarely parameterized by distance or car occupancy. Unfortunately, in the current study, in which the travel time reliability measures were generated synthetically, the crudeness of the reliability measures prevented a more detailed parameterization of VOR.

Several further research directions became very clear. Probably the most important behavioral observation is that VOT or VOR are inherently entire-day measures rather than trip-level measures. It is impossible to understand the travel behavior and choices of an individual by analyzing one particular trip taken out of the entire-day context. Daily schedules and associated time pressures, as well as monetary constraints, result in trade-offs across different activities and trips. The team believes that the most important direction of analysis should be associated with entire individual daily patterns. In this regard, discrete choice modeling techniques should be complemented by microeconomic techniques.

New types and dimensions of analysis should be supported by new types of data. It is very important to start collecting data related to schedule constraints, at least for work activity in RP travel surveys. The stereotype of a worker who has to be exactly on a fixed time schedule to and from work every day is becoming less relevant with the growing share of workers with flexible schedules. According to the latest household surveys in such major metropolitan areas as Chicago, Illinois, and San Francisco, California, less than one-quarter of workers

have a fixed schedule; more than three-quarters have at least some flexibility. Schedule flexibility logically proved to be strongly correlated with worker's income. In this regard, with better data the team could have better separated VOT-related and VOR-related effects, which looked closely correlated with the data available for the current study. For example, it is well-known that high-income workers should have a higher VOT than low-income workers (and consequently a higher willingness to pay for a toll road that reduces average travel time). In contrast, low-income workers might have a higher VOR than high-income workers because of the lower-income workers' more rigid schedule constraints.

Avoiding Simplistic Approaches to Forecasting

Key Finding

Although a number of key effects and tendencies related to the highway utility function have been tested and found to be similar across data sets and regions in the United States, many additional effects associated with person types, household composition, transit availability, and land use vary and are specific in each region. Therefore, any simplified surrogate equations or elasticity calculations need to be interpreted and applied with a great deal of caution.

Implications for Policy

Interregional comparisons and analogies and general rules with respect to expected demand elasticity to congestion and pricing have to be cautiously applied. In general, they should not be used for the evaluation of pricing projects and policies or for comparisons among different pricing alternatives. In the team's view, the importance of properly portraying congestion and pricing effects, as well as the large magnitude of possible impacts (positive or negative), fully justifies a serious modeling approach with a corresponding data collection effort. In general, the best modeling framework for congestion-related and pricing studies is a complete regional travel model system in which an advanced travel demand model is integrated with an advanced network simulation tool.

Implications for Modeling

The functional forms for the highway utility function developed in the SHRP 2 C04 research should be applied within a framework of regional travel models in which all needed structural inputs and market segments can be supported. In each particular region, the travel model can fully address regional specifics, as well as take advantage of the available data. The best framework is a complete regional travel model system in which an advanced travel demand model (preferably an activity-based microsimulation type) is integrated with an advanced network simulation tool (preferably DTA with microsimulation of individual vehicles). Analysts looking for guidance on how to capture more detail in modeling should refer to the models in the Bibliography's sources and the Appendix A.

Related Technical Detail

The findings regarding the form of the highway utility function above have been gleaned from behavioral models that have hundreds of different parameters, so any conclusions based solely on these always run the risk of ignoring some variables that might be important in specific contexts. For example, the same highway utility function can perform in a different way in a different mode choice context. If transit service is competitive, congestion pricing can result in a significant reduction of highway congestion due to the modal shift to transit. However, if transit service is not attractive, highway demand might be very inelastic with respect to congestion pricing. In the same vein, differences in income might result in different responses to congestion pricing. Low-income workers might form carpools more frequently, which would result in only partial congestion relief. Medium- and high-income workers would switch mostly to transit (especially if commuter rail is available and a park-and-ride option is convenient). In terms of congestion relief, the demand might be more elastic with respect to higher-income workers than low-income workers. In any case, pricing is not the only factor, and it is not an absolute factor defining travelers' responses. As embedded in the highway utility function, pricing works in combination with average travel time savings and reliability improvements. The trade-off between these multiple factors defines the travelers' responses. However, travel time savings and reliability improvements can only be estimated in a framework of a complete travel model, and these estimates can be very different for different O-D pairs.

This means that the developed functional forms for the highway utility function should be applied in the framework of regional travel models with all needed structural inputs and market segments. Applying these functions without the context of structural inputs and in a simplified way may result in significant aggregation biases. In the same way, operating with crude average elasticities or transferring some observed or modeled elasticities from region to region can be misleading.

It is probably impossible to develop a single universal and fully transferable model that would perform in each region equally well. The function forms developed in the current study can be used as a basic model, and they proved to be generic across different regions, including New York and Seattle. However, each regional travel model has to be designed, estimated, and calibrated to meet regional conditions.

Data Limitations and Global Positioning System–Based Data Collection Methods

Key Finding

The availability of data sets adequate to support the analyses undertaken in this study was extremely limited, especially for the aspect of travel time reliability. The culture and methodology for collecting needed travel time variability measures with O-D travel time trajectory data (not just link-level data) on a routine basis is still in its infancy, although the use of global positioning system (GPS) and probe vehicle data and other distributed wireless technologies to collect data on actual travel times and speeds is growing rapidly.

Implications for Policy

With the arrival of more comprehensive and credible data on travel times and speeds, including measures of travel time reliability, policy makers will have a significantly better basis for advocating new projects and policies, including pricing. The entire issue of improving travel time reliability can finally be transferred from the realm of qualitative analysis ("we can significantly reduce travel delays") to the quantitative analysis domain ("we can eliminate 10 occurrences a year of delays over 60 minutes"). In this regard, it is important to consider the experience of countries such as France, the Netherlands, the United Kingdom, and Japan, where improvement of travel time reliability has already been included in recommended methods for user benefit evaluation (economic appraisal) for highway projects.

Implications for Modeling

New data on travel times can form a much better basis for estimation and calibration of travel demand models and network simulation tools. Crude LOS variables created by static assignment procedures have always been one of the weakest components in travel modeling, frequently manifested in illogical values of model coefficients that need to be constrained in order to ensure reasonable model sensitivities to the network improvements. All travel demand and network simulation models would benefit from better estimates of O-D travel times by TOD. Special benefits would be provided to and could be exploited by advanced models that incorporate travel time reliability measures, such as the models developed in the current study. These new sources of information are essential for analysis and estimation of the impact of travel time reliability on travel demand.

Related Technical Detail

It is crucial in future research to take advantage of new data sources, and in particular data on travel time reliability (travel time distributions), which is currently being investigated in SHRP 2 L04, Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. As the team quickly recognized, neither of the existing RP surveys included any data on travel time reliability. A special method for generating synthetic reliability skims (i.e., O-D travel time distributions) was developed and applied to produce reliability measures for the New York and Seattle regions. However, this method had its limitations and represents only a crude surrogate for real-world travel time variation. In particular, this method cannot fully address nonrecurrent sources of congestion (like traffic incidents). At present, a growing number of principally new sources of information on highway times are becoming available. For travel demand modeling, the most important type of information is a distribution of O-D travel times for the same hour across multiple days (ideally all days of the year). With the new sources of information, such as GPS-based individual vehicle trajectories in time and space, this type of database can be built and maintained at the regional level. The team believes that using actual travel times and travel time distributions instead of synthetic skims may reveal additional important details about travelers' perception of reliability.

As the proposed reliability evaluation framework is based on travel times reported or estimated on a per vehicle trajectory basis, the travel time data required to support this research need to satisfy the following requirements:

- Report travel times by vehicle trip on a trajectory basis; at a minimum, provide *x–y* coordinates and time stamp at each reported location;
- Capture both recurring and nonrecurring congestion on a range of road facilities (from freeways to arterial roads and possibly managed lanes);
- Represent sufficient sampling and time-series to allow statistically meaningful analysis; and
- Provide the ability to tie travel time data to other ancillary or support data for time variability sources (to allow parameterization for simulation testing purposes).

The emergence of probe data over the past few years has opened the opportunity to capture all necessary data for this type of research, because these systems provide data all the time for all major roads in the network, including major arterials. The detail in such systems makes it possible to analyze travel time data according to network and route components and geographic aggregations (O-D).

In the not-too-distant past, probe-based travel times were primarily available through public- or private-sector commercial vehicle fleets (e.g., trucks, taxis, and transit vehicles) equipped with GPS technology. Travel times reported by such probe vehicles are not always fully representative of traffic

conditions, nor are the trajectories from buses and taxis particularly useful for analysis purposes. With the proliferation and wide-area penetration of wireless (cellular) telephony a few years ago, new technologies were developed to capture the location of moving cell phones and monitor them (anonymously) for purposes of gathering and analyzing travel time data. Companies around the world (e.g., ITIS, AirSage, Globis, Cel-Loc, and Intellione) developed their own technologies, such as systems and algorithms for filtering data, map matching, and time estimation, but accuracy due to technology considerations and institutional issues have not allowed some of these systems to achieve widespread deployment and use.

In recent years, GPS-based in-vehicle navigation has matured into a rapidly growing industry, and its penetration rate in the United States is probably already over 10%. A new generation of commercial in-car navigation system successfully developed and deployed by Dash Navigation, Inc. provides two-way connectivity through a built-in Wi-Fi or cellular connection that allows a network of equipped drivers to obtain up-to-date traffic flow information, make smart-route decisions, and anonymously share their speeds and locations. These GPS probe data are also recorded and available for later use in evaluating or optimizing transportation system performance. Similarly, other makers of personal navigation devices, such as TomTom, Garmin, Mio, and Magellan, plan to launch or have already launched similar Internet-connected GPS navigation systems. Many smart phone manufactures, led by Nokia, are also promoting the use of GPS-enabled mobile devices to share traffic probe data and provide collaborative location-based services with public sectors. The Mobile Century and Mobile Millennium projects in the Bay area are examples of Nokia's initiative.

Although still early in their development, some probe data systems have moved beyond the era of pure experimentation and have already evolved into full commercial applications. Systems exist based on cell phone location, GPS-equipped vehicles (usually using fleet management systems for trucks, taxis, and other commercial vehicles), and cell phones or personal digital assistants with GPS systems. In the past couple of years, the pace of the deployment of these systems has increased, including:

- National GPS-based system in the United States (operated by Inrix);
- National cellular-based system in the United States (operated by AirSage and Sprint);
- Regional (and soon to be national) cellular-based system in Canada (operated by Intellione and Rogers);
- Regional GPS- and cellular-based system (operated in Missouri by ITIS and Delcan);
- Regional (and soon to be national) truck-based GPS system that includes data on origins, destinations, and routes (operated by Calmar Telematics);
- Regional (Northern California) system based on GPS-equipped cell phones (operated by Nokia); and
- Regional systems based on Bluetooth detection.

Probe data represent a significant increase in the quality and quantity of traffic data. To realize the full value of these data requires the ability to integrate them with more traditional sources of information, and frequently both probe and wireless sources are mixed with traditional loop detection sources.

Network Simulation Models to Support Congestion and Pricing Studies

This research project addressed recent advances in traffic microsimulation tools, dynamic equilibrium algorithms, and implementation techniques for large-scale network applications, richer behavior representation in network models, and ways to generate travel time distributions and reliability measures. The results of the current study with respect to the network simulation tools are presented in Chapter 5 in detail. Salient points of the research include the following:

- Need for Microsimulation. Capturing user responses to pricing and reliability is best accomplished through microsimulation of individual traveler decisions in a network platform. These responses must be considered in a network setting, not at the facility level, and the time dimension is essential to evaluating the impact of congestion pricing and related measures. Hence a time-dependent analysis tool is required. Microsimulation of individual traveler choices provides the most general and scalable approach to evaluate the measures of interest in this study.
- More Robust DTA Required. Simulation-based DTA models have gained considerable acceptance in the past few years, yet adoption in practice remains in its infancy. The current generation of available models only considers fixed, albeit time-varying, O-D trip patterns. Greater use and utility will result from consideration of a more complete set of travel choice dimensions by integrating DTA with an activity-based demand model and incorporating user attributes, including systematic and random heterogeneity of user preferences.
- Improved Algorithms for Regional Scale Modeling. In the past, finding equilibrium time-varying flows has been based on the relatively inefficient method of successive averages, the implementation of which in a flow-based procedure did not scale well for application to large metropolitan networks. New implementations of the method of successive averages and other algorithms that exploit the vehicle-based approach of simulation-based DTA have

been proposed and demonstrated on large actual networks in this research effort.

- Traveler Heterogeneity. One of the most important conclusions of the SHRP 2 C04 project is that incorporating heterogeneity of user preferences is an essential requirement for modeling user responses to pricing in both travel demand models and network simulation tools. New algorithms that exploit nonparametric multicriteria shortest-path procedures allow VOT (which determines users' choice of path and mode in response to prices) to be continuously distributed across users. Efficient implementations of these algorithms have been demonstrated for large network application as part of this study.
- Network Reliability Measures. Most simulation models do not produce reliability estimates of travel time along network links and paths. In particular, a network simulation model has to meet two requirements: (1) route choice has to include reliability measures in a way consistent with mode choice and other choices, and (2) network path–building algorithms must generate the necessary O-D measures to feed back to the demand model along with average travel time and cost. Two practical approaches have been proposed as part of this work to estimate variability measures of travel time in the context of network assignment tools. The first exploits trajectory information in micro- and mesosimulation tools; the second employs a robust relation established between the first and second moments of the travel time per unit distance. These approaches are illustrated for application in conjunction with network evaluation tools. These methods are fully compatible with the adopted functional form of the highway utility and reliability measures, such as standard deviation of travel time per unit distance. Several new directions are currently being explored in SHRP 2 L04, Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. They include multiple network simulations (scenarios), establishing a statistical linkage between the average level of congestion and expected variability of travel times, and incorporating schedule delay penalties in a joint route and departure time choice. The team's opinion is that almost every one of the identified directions can justify a substantial research project in itself. In particular, an explicit modeling of travel time variability through managing demand and network scenarios could be of great practical value.

The proposed integrated model framework is a demonstration of a trip-based integration of a well-calibrated mode choice model in practice and a simulation-based dynamic traffic microassignment model. However, this framework is also sufficiently flexible to incorporate other dimensions (e.g., destination choice and departure time choice) in addition to the mode choice dimension from the demand side. In addition, this framework can be readily extended to an activity-based integration of demand models and an activity-based dynamic traffic microassignment model. The team believes that this study provides the theoretically and methodologically sound and complete approach needed to address heterogeneous user responses to congestion, pricing, and reliability in large-scale regional multimodal transportation networks.

This report presents the dynamic mode share and toll road usage results of the proposed integrated model on the large-scale New York metropolitan network. These results demonstrate that the model can be used on practical large-scale networks. The team also examined the convergence of the proposed algorithms. The proposed model, together with the implementation techniques described in this report, uniquely address the needs of metropolitan areas and agencies for prediction of mode and path choices and the resulting network flow patterns and provide the capability of evaluating a range of road-pricing scenarios on a large-scale network.

Incorporation of Results in Applied Travel Models

It is important to ensure that the results of the current and subsequent research are applicable within the framework of an operational travel model. Different model structures offer different options for the inclusion of advanced forms of the highway utility function. Although certain components can be incorporated in any properly segmented model, others, like travel time reliability measures or probabilistically distributed VOT, impose strict constraints on the model structure. The main related issues of incorporation of the proposed form of utility function are addressed for travel demand models and network simulation tools in the following subsections.

Transferability of Model Structures and Parameters Between Regions, Choice Contexts, and Studies

The first issue relates to the very notion of which findings and products of the C04 research can be incorporated into modeling practice. The results of the current study should be understood at three levels of generalization: (1) understanding of general rules of travel behavior and identification of major impacts and mechanisms leading to conceptual model structures, (2) understanding of mathematical structures of associated choice models and associated forms of the highway utility function, and (3) understanding of estimated choice models with the obtained values of coefficients and significance of particular variables. Which C04 findings and products can be used for other studies, and under what circumstances can they be used?

The current research has shown that at the first two levels of transferability, the model approaches and structures can be effectively generalized. Most of the functional forms for highway utility proved to be statistically significant in such different regions as New York City and Seattle. There was also a good deal of agreement between major findings based on the analysis of both RP and SP types of data. What should be undertaken with caution, however, is a direct transfer of model coefficient values from region to region, or from choice context to choice context. For different areas, even very similar choice contexts such as trip departure time versus tour departure–arrival combination, or trip mode choice versus tour mode choice, may require a significant rescaling of parameters. In practice, it also may be difficult to ensure exactly the same level of model segmentation and definition of all person, household, zonal, and LOS variables as those used in the current study.

The best way to transfer a model structure from region to region, or setting to setting, is to reestimate the model based on local data using the model specification in the current study as the prototype. This is not a simple task, but it is not nearly as complicated as model estimation from scratch, because all the structural features and variables have already been identified. In the transferability tests (e.g., from New York to Seattle and vice versa) in the present study, the absolute majority of model coefficients that had proven to be significant for one region were significant for the other region; however, the values proved to be somewhat different.

A second-best approach, which can be adopted in practice, is to recalibrate the model on aggregate local data rather than fully reestimating it in a disaggregate fashion. Recalibration can be done after the model has been implemented and the results have been compared with the aggregate targets externally established for each choice dimension. The major difference between recalibration and full reestimation is that only a subset of model parameters (bias constants that do not interact with any person, household, land use, or LOS variables) is allowed to change. In route choice, there can be only one constant (i.e., toll-averse bias). In mode choice, there is a full set of mode-specific constants. In trip departure-time choice, there are departure-specific constants (baseline departure profile) for each 30 or 60 minutes depending on the temporal resolution. In tour TOD choice there are three sets of constants depending on the model specification; for example, (1) departure-from-home profile, (2) arrival-back-home profile, and (3) activity duration profile.

Using Study Results in Applied Forecasting Models

An applied model in forecasting has to meet certain requirements that in turn impose some objective limitations on the functional forms of highway utility, and specifically on travel time reliability measures. According to the adopted levels of sophistication, the research results of this study are grounded in one or more of four applied modeling contexts:

- Aggregate (Four-Step) Demand Models. In general, these models offer a very limited framework for the incorporation of congestion and pricing effects. However, some of the main features of the suggested form of the highway utility function can be incorporated. The most constructive way to implement this is to include the suggested generalized cost components in the mode choice utilities for highway modes. The mode choice model has to differentiate highway modes by three to four occupancy categories and toll versus nontoll route type, which would result in six to eight highway modes. The model has to be segmented by trip purpose (at least two purposes, work and nonwork) and by four to five income groups to create a reasonable income distribution effect. In combination with four to five TOD periods to support a reasonable segmentation of the LOS variables, this level of segmentation may result in several hundred trip tables to manipulate. However, these technical difficulties can be overcome. The problem is that this would represent a dead-end approach because any additional segmentation by person, household, or land use characteristics or adding additional choice models (e.g., TOD choice or peak spreading) would be impossible.
- ABMs Implemented in a Microsimulation Fashion. These models are characterized by a fully disaggregate structure and rely on individual microsimulation of households and persons. They take full advantage of a detailed level of segmentation by household and person characteristics and can include complicated decision-making chains and behavioral mechanisms. The suggested form of the highway utility can be fully implemented, including route type choice, mode choice, and TOD choice as described in detail in Chapter 4. Such important variables as income and parameters like VOT can be continuously distributed to account for unobserved heterogeneity (situational variation).
- Static Traffic Assignment. It is probably impossible to incorporate travel time reliability measures in this framework except by use of simplified proxies. However, the team formulated several simplified approaches that can be implemented with these models, since in current practice they are still in use by many metropolitan planning organizations and departments of transportation. For example, the perceived highway time concept can be readily incorporated on both the demand and network simulation sides. Some improvements to the current state of the practice can be achieved with a multiclass assignment in which vehicle classes are defined by occupancy, route type (toll versus nontoll users), and (possibly) VOT-based groups (high VOT versus low VOT). However, this may result in

more than 20 vehicle classes and long run times for large regional networks.

- DTA with Microsimulation of Individual Vehicles. These models are characterized by a fully disaggregate structure and rely on individual microsimulation of vehicles. Similar to ABMs, they can take full advantage of a detailed level of segmentation by household and person characteristics linked to each vehicle, and they can also incorporate probabilistically distributed VOT in order to account for unobserved user heterogeneity. With the new technical features described in Chapter 5, these models can incorporate the suggested O-D measures of travel time reliability such as standard deviation of travel time per unit distance in the route choice, as well as generate reliability skims to feed back to the demand model.

The major applications framework for the proposed models primarily considers the full regional model framework, although facility- or corridor-level models are also considered. It is based on the recognition that for a deep understanding and proper modeling of congestion and pricing impacts, a full framework is needed. That is, a full regional travel data set and model with chosen and nonchosen alternatives available to both users and nonusers is needed. At both the model estimation stage and the application stage, it is essential to know LOS variables such as travel time, cost, and reliability for nonchoice routes, modes, TOD periods, and destinations. This necessary holistic framework is generally missing in simplified models and surveys, which limits their utility.

In general, all four combinations of the two demand model types by the two network simulation approaches are technically possible. However, it should be noted that eventually the quality of the entire travel model system is not an average of the quality of the demand and network parts, but rather reflects their minimum. That is, the weakest component, with its aggregation biases and other limitations, defines the level of resolution and accuracy of the overall modeling system. In this sense, the most promising long-term direction is for an integration of activity-based demand model with DTA, in which both models are implemented in a fully disaggregate microsimulation fashion, with an enhanced typological, temporal, and spatial resolution.

Incorporation of Travel Time Reliability in Operational Models

The incorporation of travel time reliability measures in demand models, and especially in network simulation models, still represents a major challenge, especially if the modeling system is to be practical in terms of run time and data support. Travel time reliability played a prominent role in this research, and the team implemented an extensive review and assessment of all existing approaches in terms of theoretical consistency and applicability in operational models, as reported in detail in Chapter 4. In general, four possible methodological approaches to quantifying reliability are either suggested in the research literature or already applied in operational models:

- Indirect Measure: Perceived Highway Time by Congestion Levels. This concept is based on statistical evidence that in congestion conditions, travelers perceive each minute with a certain weight. Perceived highway time is not a direct measure of reliability because only the average travel time is considered, although it is segmented by congestion levels. It can, however, serve as a good instrumental proxy for reliability because the perceived weight of each minute spent in congestion is a consequence of associated unreliability.
- First Direct Measure: Time Variability (Distribution) Measures. This is considered as the most practical direct approach and has received considerable attention in recent years. This approach assumes that several independent measurements of travel time are known that allow for forming the travel time distribution and calculation of derived measures, such as buffer time. An important technical detail with respect to generation of travel time distributions is that even if the link-level time variations are known, it is a nontrivial task to synthesize the O-D–level time distribution (reliability skims) because of the dependence of travel times across adjacent links due to mutual traffic flow.
- Second Direct Measure: Schedule Delay Cost. This approach has been adopted in many academic research works on individual behavior. According to this concept, the direct impact of travel time unreliability is measured through cost functions (penalties in expressed in monetary terms) of being late (or early) compared with the planned schedule of the activity. This approach assumes that the desired schedule is known for each person and activity undertaken in the course of the modeled period. This assumption, however, is difficult to meet in a practical model setting.
- Third Direct Measure: Loss of Activity Participation Utility. This method can be thought of as a generalization of the schedule delay concept. It is assumed that each activity has a certain temporal utility profile and that individuals plan their schedules to achieve maximum total utility over the modeled period (e.g., the entire day), taking into account expected (average) travel times. Any deviation from the expected travel time due to unreliability can be associated with a loss of participation in the corresponding activity (or gain if travel time proved to be shorter). Recently this approach was adopted in several research works on DTA formulation integrated with activity scheduling analysis. Similar to the schedule delay concept, however, this approach suffers from data requirements that are difficult to meet in practice. The added complexity of estimation

and calibration of all temporal utility profiles for all possible activities and all person types is significant. This complexity made it unrealistic to adopt this approach as the main concept for the current project. This approach, however, can be considered in future research efforts.

Current possibilities for incorporating each approach in operational models, supported with the necessary input data, are summarized in Table 6.2. The main aspects are analyzed within the specific frameworks of demand modeling and network simulation. Both sides are equally important in order to construct a complete operational regional model.

In summary, as a proxy for reliability, perceived highway time can be easily incorporated in travel models in the short term because it does not require any significant restructuring of either the demand or network simulation side. It can be implemented with a traditional four-step demand model combined with a simple static assignment procedure. However, this can be only a temporary solution, because it is not a true incorporation of travel time reliability.

Reliability measures based on travel time distribution, with measures such as variance or standard deviation, can be relatively easily incorporated in travel demand models as an additional variable. The models themselves do not need to be significantly restructured compared with the existing advanced structures already applied in ABMs. A bigger challenge is to support travel time reliability on the network simulation side, which can only be done with an advanced network simulation model of the DTA type. This model also needs to be route based (rather than link based), which imposes additional computational challenges. This direction and approach were adopted for the current study.

More advanced and theoretically appealing approaches that are based on schedule delay cost and loss in activity participation are more problematic to implement. There are some significant challenges on the demand side, such as the specification of preferred arrival times for all trips, which are generally not known in RP surveys. However, probably a bigger challenge is to develop a network simulation tool that could generate realistic O-D travel time distributions instead of predetermined scalar measures like variance or standard deviation. Thus, these approaches are only suggested for future research.

Further investigation into more advanced reliability measures and ways to incorporate them in travel demand models and network simulation tools is currently being implemented in SHRP 2 L04, Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools.

Summary of Recommended Model Parameters

A summary of the recommended (default) values for all coefficients applied in the highway utility function is provided in Table 6.3. These parameters have been established based on the models statistically estimated in the SHRP 2 C04 research, as well as derived from other comparable models reported in literature. These parameters are recommended for use in operational models only if a full disaggregate estimation of the regional data cannot be implemented. In that case, a careful aggregate validation and calibration of the entire model system, including route type choice, mode choice, and TOD choice, will be needed.

Recommendations for Future Research

General Considerations

The current project represented a unique opportunity to explore a wide range of effects associated with congestion

(text continues on page 156)

Table 6.2. Incorporation of Travel Time Reliability in Operational Models

Method	Demand Model	Network Simulation
Perceived highway time	Straightforward and does not require structural changes	Straightforward and does not require structural changes
Time distribution (mean variance)	Straightforward and does not require structural changes	Network route choice has to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D reliability measures need to be generated
Schedule delay cost	Preferred arrival time has to be externally specified for each trip	Network route choice has to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D travel time distributions should be generated either analytically or through multiple simulations
Loss of participation in activities	Temporal utility profiles have to be specified for each activity; entire-day schedule consolidation model has to be applied	Network route choice needs to incorporate reliability measures that are not additive by links; this requires explicit route enumeration. O-D travel time distributions have to be generated either analytically or through multiple simulations

Table 6.3. Recommended Coefficient Values

	Model Coefficients								Examples of Population and Travel Characteristics			Derived Measures					
			Distance (mi)														
Travel Purpose	Toll Bias	Time (min)	Linear	Squared	Cost (cents)	SD per mi (min/mi)	Exponent for Income	Exponent for Car Occupancy	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Toll Bias Equivalent (min)	Cost Coefficient with Income and Occupancy Effects	VOT ($/h)	VOR ($/h)	Reliability Ratio
To Work and Business	–0.85	–0.0425	0.02024	–0.000266	–1.25	–0.625	0.6	0.8	30,000	1.0	5.0	–0.0465	18.3	–0.0026	10.8	29.1	2.69
									30,000	2.0	5.0	–0.0465	18.3	–0.0015	18.9	50.7	2.69
									30,000	3.0	5.0	–0.0465	18.3	–0.0011	26.1	70.2	2.69
									30,000	1.0	10.0	–0.0500	17.0	–0.0026	11.6	14.6	1.25
									30,000	2.0	10.0	–0.0500	17.0	–0.0015	20.3	25.4	1.25
									30,000	3.0	10.0	–0.0500	17.0	–0.0011	28.0	35.1	1.25
									30,000	1.0	20.0	–0.0552	15.4	–0.0026	12.9	7.3	0.57
									30,000	2.0	20.0	–0.0552	15.4	–0.0015	22.4	12.7	0.57
									30,000	3.0	20.0	–0.0552	15.4	–0.0011	31.0	17.5	0.57
									60,000	1.0	5.0	–0.0465	18.3	–0.0017	16.4	44.2	2.69
									60,000	2.0	5.0	–0.0465	18.3	–0.0010	28.6	76.9	2.69
									60,000	3.0	5.0	–0.0465	18.3	–0.0007	39.6	106.4	2.69
									60,000	1.0	10.0	–0.0500	17.0	–0.0017	17.7	22.1	1.25
									60,000	2.0	10.0	–0.0500	17.0	–0.0010	30.7	38.4	1.25
									60,000	3.0	10.0	–0.0500	17.0	–0.0007	42.5	53.2	1.25
									60,000	1.0	20.0	–0.0552	15.4	–0.0017	19.5	11.0	0.57
									60,000	2.0	20.0	–0.0552	15.4	–0.0010	33.9	19.2	0.57
									60,000	3.0	20.0	–0.0552	15.4	–0.0007	46.9	26.6	0.57
									100,000	1.0	5.0	–0.0465	18.3	–0.0013	22.3	60.0	2.69
									100,000	2.0	5.0	–0.0465	18.3	–0.0007	38.9	104.5	2.69
									100,000	3.0	5.0	–0.0465	18.3	–0.0005	53.8	144.5	2.69
									100,000	1.0	10.0	–0.0500	17.0	–0.0013	24.0	30.0	1.25
									100,000	2.0	10.0	–0.0500	17.0	–0.0007	41.8	52.2	1.25
									100,000	3.0	10.0	–0.0500	17.0	–0.0005	57.8	72.2	1.25
									100,000	1.0	20.0	–0.0552	15.4	–0.0013	26.5	15.0	0.57
									100,000	2.0	20.0	–0.0552	15.4	–0.0007	46.1	26.1	0.57
									100,000	3.0	20.0	–0.0552	15.4	–0.0005	63.8	36.1	0.57
From Work and Business	–0.95	–0.0425	0.02024	–0.000266	–1.44	–0.545	0.6	0.8	30,000	1.0	5.0	–0.0465	20.4	–0.0030	9.4	22.1	2.34
									30,000	2.0	5.0	–0.0465	20.4	–0.0017	16.4	38.4	2.34
									30,000	3.0	5.0	–0.0465	20.4	–0.0012	22.7	53.1	2.34
									30,000	1.0	10.0	–0.0500	19.0	–0.0030	10.1	11.0	1.09
									30,000	2.0	10.0	–0.0500	19.0	–0.0017	17.6	19.2	1.09
									30,000	3.0	10.0	–0.0500	19.0	–0.0012	24.3	26.6	1.09
									30,000	1.0	20.0	–0.0552	17.2	–0.0030	11.2	5.5	0.49
									30,000	2.0	20.0	–0.0552	17.2	–0.0017	19.4	9.6	0.49
									30,000	3.0	20.0	–0.0552	17.2	–0.0012	26.9	13.3	0.49
									60,000	1.0	5.0	–0.0465	20.4	–0.0020	14.3	33.4	2.34
									60,000	2.0	5.0	–0.0465	20.4	–0.0011	24.8	58.2	2.34
									60,000	3.0	5.0	–0.0465	20.4	–0.0008	34.4	80.5	2.34
									60,000	1.0	10.0	–0.0500	19.0	–0.0020	15.3	16.7	1.09
									60,000	2.0	10.0	–0.0500	19.0	–0.0011	26.7	29.1	1.09

(continued on next page)

Table 6.3. Recommended Coefficient Values (continued)

	Model Coefficients								Examples of Population and Travel Characteristics			Derived Measures					
			Distance (mi)														
Travel Purpose	Toll Bias	Time (min)	Linear	Squared	Cost (cents)	SD per mi (min/mi)	Exponent for Income	Exponent for Car Occupancy	Household Income ($/year)	Car Occupancy	Distance (mi)	Time Coefficient with Distance Effect	Toll Bias Equivalent (min)	Cost Coefficient with Income and Occupancy Effects	VOT ($/h)	VOR ($/h)	Reliability Ratio
From Work and Business, continued									60,000	3.0	10.0	–0.0500	19.0	–0.0008	36.9	40.3	1.09
									60,000	1.0	20.0	–0.0552	17.2	–0.0020	16.9	8.4	0.49
									60,000	2.0	20.0	–0.0552	17.2	–0.0011	29.5	14.6	0.49
									60,000	3.0	20.0	–0.0552	17.2	–0.0008	40.8	20.1	0.49
									100,000	1.0	5.0	–0.0465	20.4	–0.0014	19.4	45.4	2.34
									100,000	2.0	5.0	–0.0465	20.4	–0.0008	33.7	79.1	2.34
									100,000	3.0	5.0	–0.0465	20.4	–0.0006	46.7	109.4	2.34
									100,000	1.0	10.0	–0.0500	19.0	–0.0014	20.8	22.7	1.09
									100,000	2.0	10.0	–0.0500	19.0	–0.0008	36.3	39.5	1.09
									100,000	3.0	10.0	–0.0500	19.0	–0.0006	50.1	54.7	1.09
									100,000	1.0	20.0	–0.0552	17.2	–0.0014	23.0	11.4	0.49
									100,000	2.0	20.0	–0.0552	17.2	–0.0008	40.0	19.8	0.49
									100,000	3.0	20.0	–0.0552	17.2	–0.0006	55.4	27.3	0.49
Nonwork	–1.2	–0.0335	0	0	–0.5228	–0.418	0.5	0.7	30,000	1.0	5.0	–0.0335	35.8	–0.0030	6.7	16.6	2.50
									30,000	2.0	5.0	–0.0335	35.8	–0.0019	10.8	27.0	2.50
									30,000	3.0	5.0	–0.0335	35.8	–0.0014	14.4	35.9	2.50
									30,000	1.0	10.0	–0.0335	35.8	–0.0030	6.7	8.3	1.25
									30,000	2.0	10.0	–0.0335	35.8	–0.0019	10.8	13.5	1.25
									30,000	3.0	10.0	–0.0335	35.8	–0.0014	14.4	17.9	1.25
									30,000	1.0	20.0	–0.0335	35.8	–0.0030	6.7	4.2	0.62
									30,000	2.0	20.0	–0.0335	35.8	–0.0019	10.8	6.7	0.62
									30,000	3.0	20.0	–0.0335	35.8	–0.0014	14.4	9.0	0.62
									60,000	1.0	5.0	–0.0335	35.8	–0.0021	9.4	23.5	2.50
									60,000	2.0	5.0	–0.0335	35.8	–0.0013	15.3	38.2	2.50
									60,000	3.0	5.0	–0.0335	35.8	–0.0010	20.3	50.7	2.50
									60,000	1.0	10.0	–0.0335	35.8	–0.0021	9.4	11.8	1.25
									60,000	2.0	10.0	–0.0335	35.8	–0.0013	15.3	19.1	1.25
									60,000	3.0	10.0	–0.0335	35.8	–0.0010	20.3	25.4	1.25
									60,000	1.0	20.0	–0.0335	35.8	–0.0021	9.4	5.9	0.62
									60,000	2.0	20.0	–0.0335	35.8	–0.0013	15.3	9.5	0.62
									60,000	3.0	20.0	–0.0335	35.8	–0.0010	20.3	12.7	0.62
									100,000	1.0	5.0	–0.0335	35.8	–0.0017	12.2	30.3	2.50
									100,000	2.0	5.0	–0.0335	35.8	–0.0010	19.8	49.3	2.50
									100,000	3.0	5.0	–0.0335	35.8	–0.0008	26.2	65.5	2.50
									100,000	1.0	10.0	–0.0335	35.8	–0.0017	12.2	15.2	1.25
									100,000	2.0	10.0	–0.0335	35.8	–0.0010	19.8	24.6	1.25
									100,000	3.0	10.0	–0.0335	35.8	–0.0008	26.2	32.7	1.25
									100,000	1.0	20.0	–0.0335	35.8	–0.0017	12.2	7.6	0.62
									100,000	2.0	20.0	–0.0335	35.8	–0.0010	19.8	12.3	0.62
									100,000	3.0	20.0	–0.0335	35.8	–0.0008	26.2	16.4	0.62

Note: SD = Standard deviation.

(continued from page 151)
and pricing on travel demand in a comprehensive framework of various travel dimensions including auto route choice, mode choice, and TOD choice. However, like any research, it had to be limited to a finite number of model components and bound to available data sets. In the process of statistical analysis and behavioral interpretations of the results, many additional ideas were generated, and possible directions for model improvements were identified that could not be fully addressed in the current project. The team summarizes its recommendations for future research along the following directions:

- Take advantage of new data sources, and in particular data on travel time reliability (travel time distributions) as it is currently being investigated in SHRP 2 L04, Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. As the team quickly recognized, neither of the existing RP surveys had included any data on travel time reliability. A special method for generating synthetic reliability skims (i.e., O-D travel time distributions) was developed and applied to produce reliability measures for the New York and Seattle regions. However, this method had its limitations and represents only a crude surrogate for real-world travel time variation. In particular, this method cannot fully address nonrecurrent sources of congestion (like traffic incidents). At present, a growing number of principally new sources of information on highway times are becoming available. For travel demand modeling, the most important type of information is a distribution of O-D travel times for the same hour across multiple days (ideally all days of the year). With the new sources of information, such as GPS-based individual vehicle trajectories in time and space, this type of database can be built and maintained at the regional level. The team believes that using the actual travel times and travel time distributions instead of synthetic skims may reveal additional important details about travelers' perception of reliability.
- Extend the travel dimensions and choice frameworks adopted in the current study. In the current study, analysis focused on the three primary responses of highway users to congestion and pricing, which include taking a different highway route, changing the mode (e.g., switching to transit), and changing the departure time of travel (e.g., switching from the peak hour to a later hour). A general approach for incorporation of the other travel dimensions, including destination choice, trip chaining, daily activity pattern (tour and trip frequency), and car ownership, was outlined. This approach is based on using the developed models for primary choices to form a wide set of accessibility measures that can be included in all other models. This approach has some behavioral appeal in terms of the integrity of the entire model system and has been successfully applied in many ABMs in practice. However, this approach has its own limitations, and it is worth investigating if there are some direct effects of highway congestion and pricing on trip destinations, trip frequencies, car ownership, and other dimensions.
- Explore more general behavioral frameworks than a system of hierarchical discrete choice models; such exploration may include microeconomic frameworks of rational behavior under resource constraints. The econometric-based research on travel behavior has been historically dominated by discrete choice models because of the computational advantages in terms of model estimation and application. However, several aspects specifically related to highway congestion and pricing make a microeconomic framework appealing. Household and person travel is subject to time, space, and monetary constraints. It is obvious with respect to time constraints that as every person has 24 hours a day, all activities and trips have to be implemented within this constraint. It is also relatively straightforward to extend a one-dimensional time constraint to a two-dimensional time–space constraint of the individual travel patterns based on the maximum possible travel speed. A monetary constraint is the most complicated because it is fuzzier, and household and person daily budgets of different days can be traded off. However, monetary constraint also exists and strongly manifests itself in practice when an average day is modeled. A system of hierarchical choice models is awkward when dealing with these constraints because they create linkages across choices made for different trips and tours. In this sense, VOT or VOR (or both) on one trip are dependent on the other trips. A person may be willing to pay $10 for a better LOS for a particular trip if this is the only trip that uses a priced facility. However, the same person may refuse to pay $30 if he has to make three trips with similar characteristics. These satiation and bounding effects can be naturally incorporated in a microeconomic framework, but their incorporation in a discrete choice framework would frequently result in a model implosion because the corresponding choice dimensions have to be combined in a Cartesian way. Microeconomic techniques have not been widely applied in travel models because the microeconomic framework has its own limitations, primarily a high complexity in the resulting optimization problem when discreteness of trips and activities is properly accounted for. (Note: In a classic microeconomic theory of consumer behavior, the products are all continuous divisible entities). With rapidly improving computer power, it is a viable and attractive option to build

and estimate a microeconomic model of travel behavior analogous to the daily activity pattern choice model, and compare the results.

- It is important to ensure that the results of the current and subsequent research be applicable in the framework of an operational travel model. At an early stage of the project it became clear that inclusion of more sophisticated forms of highway utility (generalized cost) in travel demand models (like mode choice and TOD choice) is relatively straightforward and does not change the principal structure of these models. However, incorporation of travel time reliability measures in network simulation models still represents a big challenge, especially if the model system is to be practical in terms of run time and data support. In particular, a network simulation model has to meet two requirements: (1) route choice has to include the reliability measures in a way consistent with mode choice and other choices, and (2) network path–building algorithms have to generate the necessary O-D measures to feed back to the demand model along with average travel time and cost. Several new directions are currently being explored in SHRP 2 L04, Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. They include multiple network simulations (scenarios), establishing a statistical linkage between the average level of congestion and expected variability of travel times, and incorporation of schedule delay penalties in a joint route and departure time choice. The team's opinion is that almost every one of the identified directions can justify a substantial research project in itself. In particular, an explicit modeling of travel time variability through managing demand and network scenarios could be of great practical value.
- The team also recommends continuing research and analysis of car occupancy choices and associated carpooling mechanisms. An accurate modeling of car occupancy is essential for projects and policies involving HOV and HOT lanes. Car occupancy choice is a special choice type that is not an individual person choice. What stands behind this choice is a (frequently unseen) process of schedule synchronization between several persons. In some advanced ABMs, the first steps have been made to model some types of joint travel explicitly. From the current research, it became clear that different types of carpools may have different cost-sharing mechanisms and consequently different VOT and VOR. In particular, the type of carpool (intrahousehold versus interhousehold) and travel party composition (adults only or adults with children) are important determinants of VOT and VOR. There are also some new tendencies, such as casual carpooling in San Francisco, which have to be addressed in mode choice. In the team's view, carpooling deserves special attention and substantiates a focused research project.
- Include more intensive international comparisons and, in particular, take advantage of many interesting theoretical developments on travel time reliability in Europe. The biggest challenge would be to transfer the most interesting and theoretically consistent results from the SP realm to the RP realm. The absolute majority of European studies on travel time reliability are based on specially designed SP experiments that could not have been replicated in an RP setting. The primary obstacle was the absence of the observed data on travel time distributions for the needed O-D trips. A potential breakthrough can happen if this direction is combined with the previously discussed general direction on using new sources of information on highway travel times.

Implementation Opportunities

This C04 research project, Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand, is one of the first to bridge the objectives of both the Capacity and the Reliability research programs at SHRP 2. Major highway or system interventions (e.g., highway supply and operations actions, travel demand management, pricing, information, and new technologies) directly and indirectly affect both the delivered capacity (throughput) and reliability of the service, as shown in the framework of Figure 6.6.

The project addressed the connections between capacity, congestion, and reliability through user responses to interventions (pricing) and service levels (congestion and reliability). However, as noted previously, the user response models developed in this project were limited by the availability of reliability information (supply-side attributes). Nonetheless, the framework elaborated here for integration into a network modeling platform is an important practical accomplishment.

Accordingly, the team sees three important opportunities for implementation-oriented research and additional work that would help overcome some of the limitations encountered in the study and deliver the powerful findings and tools to the practice community. These implementation opportunities are discussed in the following sections.

First Implementation Opportunity

The first opportunity would leverage new data sources to overcome the limitations of existing data encountered in the present C04 effort. Three new sources of data not available to this project may now be coming online through synergistic activities undertaken as part of SHRP 2 Project L04,

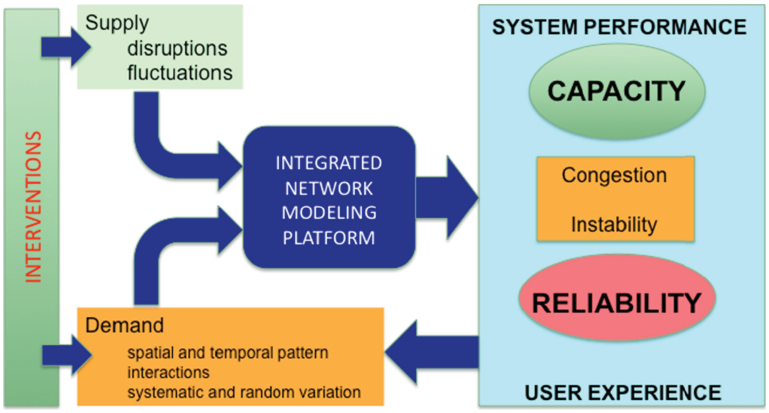

Figure 6.6. Relationships between system interventions and system performance.

Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. These data include

- A travel choices data set from Seattle that could be analyzed more extensively for supply-side variability, especially experienced variability;
- Simulated variability (objective and experienced) in the New York regional network using methods developed in L04; and
- Actual travel time probe data being acquired by several Reliability projects, including L04.

The model framework developed in the present study enables more complete behavior representation for modeling user responses to reliability; this is important for the integration efforts in SHRP 2 Project C10, Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network. Additional effort as proposed would also provide an opportunity to demonstrate practical models of user decisions (route, mode, and activity timing) that explicitly capture reliability. An effort on reliability measures could be useful for SHRP 2 Project L05, Incorporating Reliability Performance Measures into the Transportation Planning and Programming Processes.

Actions needed to implement the first implementation opportunity include the following:

- For project evaluation and economics, the development of improved and more realistic VOR and reliability ratios (input to L05 and other studies);
- A demonstration for the New York regional network of a working model and procedures that integrate with the network model; and
- Development of a transferable approach and model that can be used with other locations (e.g., for C10).

Second Implementation Opportunity

The second major implementation research opportunity derives from integrating improved behavior models in network modeling procedures. Such an integration would enable modeling responses to capacity improvements that affect reliability (e.g., taking the output of Project C05, Understanding the Contribution of Operations, Technology, and Design to Meeting Highway Capacity Needs).

Furthermore, through a combined demonstration of the procedures developed for Projects C04 and L04, using the already-developed and calibrated New York region model, a useful and effective blueprint framework would be obtained to support the work under Project C10. As such, it would add a demand–behavior dimension to the supply-side work envisioned under Project L04 (and currently not in the scope of that project).

Actions needed to implement the second implementation opportunity include the following:

- Development of a methodology and platform that tangibly and demonstrably integrate demand and supply-side developments in modeling reliability and capacity; and
- Actual application of the new methodology to the New York regional network, which would showcase the project results and provide an incentive for other areas (and would support Project C10).

Third Implementation Opportunity

The third opportunity consists in evaluating alternative mechanisms for incorporating reliability measures in integrated planning procedures with different situations regarding the availability of data (e.g., none, some, trajectories, single day, multiple days, and years). This opportunity builds

on the findings from Project L04 in the C04 framework to leverage analytic relations that may be used when primary data are unavailable or only partially available.

Actions needed to implement the third implementation opportunity include the following:

- Development of an application "primer" for metropolitan planning organizations and state and local agencies to explain what to do and how to do it given available modeling tools (e.g., static, dynamic, stand alone, or integrated) and available data (and the resources to collect them); and
- Provision of the currently missing methodological complement for L05, Incorporating Reliability Performance Measures into the Transportation Planning and Programming Process.

APPENDIX A

Mathematical and Procedural References

Method for Synthesizing a Distribution of Consistent Path-Dependent Origin–Destination Travel Times from the Known Distribution of Link Traffic Counts

Problem

This method is suggested for the generation of origin–destination (O-D) travel time distribution for the base year, which is needed for calculating travel time reliability measures. These reliability measures are used in travel demand models to explain travel choices along with the average travel time and cost. The current memo explains generation of O-D time distribution for model estimation for the base year. The method is designed to produce a distribution of travel times for a full regional O-D matrix for a certain time of day period or hour (i.e., period-specific "reliability skim").

The following inputs are assumed given:

$a \in A$ = highway network links;
$a \in \widetilde{A} \subset A$ = links for which traffic counts are known;
$\{w_a^n\}_{a \in \widetilde{A}, n \in N_a}$ = sets of traffic counts (discrete distribution) for each link;
$\overline{w}_a$ = average traffic count for each link;
$c_a(v_a)$ = link volume-delay function (VDF) calibrated to reproduce observed travel time distribution; and
T_{ij} = seed trip table adjusted to reproduce average traffic counts.

Without a loss of generality, we assume that a set of traffic counts for each link has the same size (say, 10 or 20 observations). If a link has fewer observations than the established number, additional counts are created by interpolation to preserve the observed distribution. For the algorithm implementation, it is convenient to order counts for each link from the smallest to the largest:

$$w_a^1 \le w_a^2 \le \cdots \le w_a^n \le \cdots \le w_a^N. \tag{A.1}$$

In general, it is assumed that traffic counts for different links are taken on different days and maybe during different seasons and even in different years. For this reason, some of them can undergo certain adjustments to bring them to a common denominator. Thus, in general, it is impossible to establish a linkage or a priori correlation pattern between them (it might be possible only for some of them taken on the same day in parallel; this partial information can be used in the proposed algorithm as an additional constraint).

The method generates the following output:

$s \in S$ = demand fluctuation scenarios, each associated by a trip table T_{ij}^s;
$x_a^{ns} = 0, 1$ = counts assortment by scenarios; and
C_{ij}^s = a set of travel time skims (discrete distribution) by scenarios.

Approach

Travel time skims by scenarios are generated by scenario-specific trip tables T_{ij}^s pivoted off the seed trip table. The scenario-specific trip tables are constructed by adjustment to scenario-specific subsets of traffic counts $\{w_a^{n(s,a)}\}_{a \in \widetilde{A}}$ where $n(s, a)$ represents a count value from the link distribution that is selected for scenario s. Scenarios can be meaningfully associated with certain demand fluctuations reflected in traffic counts (e.g., a rainy day with generally low traffic, special event with the corresponding counts taking very high values). These scenarios, however, cannot cover cases where the supply side, that is, network capacity, was reduced (like a traffic accident or lane closure because of road work).

When forming the network scenarios from link count distributions, two main principles should be followed:

- **The observed distribution of traffic counts for each link should be preserved across the network scenarios.** This can be easily achieved by constructing scenarios as combinations of permutations of the numbered counts 1, 2, . . . , N for each link. In this case, the number of scenarios is equal to the number of counts on each link (say, 10). Each count value is used once in one scenario and cannot be reused in a different scenario. Although this strategy guarantees a replication of the distribution of counts for each link, it alone is not enough to construct realistic network scenarios portraying demand fluctuations. For example, one could envision constructing simplified scenarios fully ordered by traffic counts (from the smallest to the largest, i.e., setting $s = n$). In this case, the first scenario would correspond to all lowest count values while the last scenario will correspond to all highest count values. This, however, would create an unrealistic pattern of correlations where all links in the network are assumed to fluctuate simultaneously, even though it is true only for links in the same corridor that have a substantial mutual traffic flow.
- **The correlation pattern between links across scenarios should follow the logic of traffic flows.** This means that adjacent links in the same highway corridor with a substantial mutual traffic flow (or parallel links competing for the same O-D pairs) should have a highly correlated pattern (both should have either a high or a low value in the same scenario). Contrary to that, links that do not have a substantial mutual flow and do not serve the same O-D pairs (i.e., correspond to different demand segments) should have a random correlation pattern where each link can take a value independently from the other in each scenario. In particular, one link can have a very high value and the second one can have a very low value in the same scenario. The correlation pattern is established in the assignment procedure with a trip table adjusted to traffic counts. In particular, several count values can be easily replicated by a small adjustment of the trip table if they are consistent with the flow pattern. On the other hand, if a particular count value persists not to be replicated, it is an indication of the contradiction between this count and the traffic flow pattern defined by the rest of the counts and O-D demand.

Program Formulation

We assume without essential loss of generality that the number of scenarios S is equal to the number of count values available for the links N. The problem can be stated as the following mathematical program:

$$\min_{\{T_{ij}^s>0,\, x_a^{ns}=0,1,\, v_a^s\}} \left[F_1\left(\{T_{ij}^s\}\right)+\mu_2 F_2\left(\{T_{ij}^s\}+\mu_3 F_3\left(\{x_a^{ns}, v_a^s\}\right)\right)\right], \tag{A.2}$$

where

$$F_1\left(\{T_{ij}^s\}\right)=\sum_{s,ij} T_{ij}^s \ln\frac{T_{ij}^s}{T_{ij}}, \tag{A.3}$$

$$F_2\left(\{T_{ij}^s\}\right)=\sum_{ij,kl} T_{ij}T_{kl}\rho_{ijkl}^2, \tag{A.4}$$

$$F_3\left(\{x_a^{ns}, v_a^s\}\right)=\sum_{s,a\in A}\left(v_a^s-\sum_n w_a^n x_a^{ns}\right)^2, \tag{A.5}$$

$\mu_2 > 0$, $\mu_3 > 0$ represents weights for the second and third criterion relative to the first criterion, under constraints:

$$\sum_n x_a^{ns}=1, \begin{pmatrix}\text{each scenario uses only one}\\ \text{count value for each link}\end{pmatrix} \tag{A.6}$$

$$\sum_s x_a^{ns}=1, \begin{pmatrix}\text{each count value for each link}\\ \text{is used in only one scenario}\end{pmatrix} \tag{A.7}$$

$$v_a^s=\sum_{ij} T_{ij}^s \sum_r p_{ijr}^s \delta_{ar}^s, \text{(flow conservation)} \tag{A.8}$$

where

$p_{ijr}^s \geq 0$ = path probabilities,
$\delta_{ar}^s = 0,1$ = link–path incidence matrix.

The meaning of the first criterion (Equation A.3) is that all else being equal, the scenario-specific trip tables T_{ij}^s should be structurally similar to the original/base trip table T_{ij}.

The meaning of the second criterion (Equation A.4) is that all else being equal, the scenario-specific variations of trip tables should not be systematically correlated but rather randomized across O-D pairs. In other words, we assume that demand fluctuations in different cells are independent unless the constraints and/or other two criteria indicate correlation. Randomization of the scenario-specific trip tables preserves the necessary independence of traffic flows across links except for links that have a significant mutual flow of vehicles (i.e., significant demand from the same O-D pairs). Traffic volumes on those links will always be fluctuating in parallel, even if the demand fluctuations for the corresponding O-D pairs are independent from each other. The pair-wise measure of demand correlation between O-D pairs can be written as follows:

$$\rho_{ij,kl}=\frac{\sum_s\left(T_{ij}^s-T_{ij}\right)\left(T_{kl}^s-T_{kl}\right)}{\sqrt{\left[\sum_s\left(T_{ij}^s-T_{ij}\right)^2\right]\left[\sum_s\left(T_{kl}^s-T_{kl}\right)^2\right]}}. \tag{A.9}$$

In the program formulation, this measure is squared (because independence is violated by both positive and negative correlations), weighted by the demand, and summed over all pairs of O-D cells. The original correlation coefficient is $-1 \le \rho_{ij,kl} \le 1$. The squared coefficient is in the unit interval.

It should be noted that, in general, there is no contradiction between the first and second criterion. The first criterion (Equation A.3) states that the deviations of the scenario-specific trip tables from the original average trip table should be minimal. The second criterion (Equation A.4) is not dependent on the magnitude of the deviations, but rather requests from them to be independent across O-D pairs.

The third criterion (Equation A.5) expresses the desired replication of distributions of traffic counts observed for each link.

The mathematic program formulated in Equations A.2 through A.9 is a complicated, nonlinear program of a very large size (in real-world networks) with both continuous and discrete variables. It cannot be solved by direct formal methods. This formulation serves only as a useful conceptual framework, within which a practically effective heuristic algorithm is proposed.

Heuristic Algorithm

The following algorithm is suggested to incorporate both the principles and the resulting three formal criteria (Equations A.3–A.5) described in the previous section.

Step 0. Define a seed trip table T_{ij} and adjust it to replicate average traffic counts $\{\overline{w}_a\}$. This is an auxiliary step to prepare a good starting condition. Set initially all scenario-specific trip tables to be equal to the seed trip table $T^s_{ij} = T_{ij}$.

Step 1. Randomly and independently vary each scenario-specific trip table T^s_{ij}. The variation should be implemented independently in each cell to address the second criterion (Equation A.4). The magnitude of variation is approximately equal to the average level of variation in counts. In each cell and for each scenario, an independent draw is implemented from a lognormal distribution with the mean equal to T^s_{ij} and scheduled time of departure (STD) equal to average STD observed in traffic counts. The lognormal distribution ensures positive values and avoids a truncation problem associated with the normal distribution. Assign the scenario-specific trip tables (each one separately) to obtain scenario-specific link volumes v^s_a.

Step 2. Optimize the assortment of traffic count values by scenarios in terms of x^{ns}_a and taking into account the third criterion (Equation A.5). Define x^{ns}_a in such a way that the order of traffic counts on each link would correspond to the order of link volumes. It means that for each link, scenarios are ordered as follows:

$$v_a^{s_1(a)} \le v_a^{s_2(a)} \le \cdots \le v_a^{s_n(a)} \le \cdots \le v_a^{s_N(a)}. \tag{A.10}$$

Then the assortment variables are calculated to ensure that the order of traffic counts corresponds to order of link volumes:

$$x^{ns}_a = \begin{cases} 1, & \text{if} \quad s_n(a) = n \\ 0, & \text{if} \quad s_n(a) \ne n \end{cases}. \tag{A.11}$$

After the assortment, each link obtains a scenario-specific set of counts $\{\omega^s_a\}$ by the following formula:

$$\omega^s_a = \sum_n w^n_a x^{ns}_a. \tag{A.12}$$

Step 3. Adjust the scenario-specific trip tables T^s_{ij} to the corresponding scenarios of traffic counts $\{\omega^s_a\}$. Matrix adjustment corresponds to the original program (Equations A.2–A.9) with fixed count assortment variables $\{x^{ns}_a\}$. If the assortment variables are fixed, the program is reduced to optimization by two criteria (Equations A.3, A.5) under constraint (Equation A.8). Existing effective heuristic algorithms can be employed for this sub-problem. Save the adjusted scenario-specific trip tables T^s_{ij}, traffic volumes $\{v^s_a\}$, and O-D travel times $\{C^s_{ij}\}$.

Step 4. Possible feedback can be implemented to Step 1 if the trip table adjustment after multiple iterations [i.e., compromising criterion (Equation A.3)] has not produce a good match to traffic counts. Persistent counts that cannot be matched indicate a possible inconsistency in the counts assortment by scenarios that can be re-sorted at the second iteration.

The essence of this algorithm is to reorder the traffic count values by consistent demand fluctuation scenarios where each scenario has a consistent flow pattern that replicates the chosen subset of count values. It is essential to fully exploit the switching mechanism to create consistent scenarios rather than overadjust the trip table to replicate inconsistent scenarios. For this reason, the matrix adjustment subroutine will be used with a small number of internal iterations (Equations A.2–A.3) expecting a large number of feedback iterations between Steps 1 and 3.

Implementation

The entire algorithm can be implemented using the script language of any transportation software package (Geographic Information System Developer's Kit [GISDK] for TransCAD or Macro-language for EMME). The only substantial subroutine required is a trip matrix adjustment to traffic counts available in all transportation software packages.

Mathematical Formulation of the Integrated Multidimensional Network Choice Model

This section provides additional detail on the mathematical formulation of the integrated multidimensional network travel choice model, with multimodal route and mode choice,

introduced in Chapter 4 and applied to the New York Regional Best Practice Network.

The basic model assumptions were stated in Chapter 4 within the Integrated Model Framework to Evaluate Pricing and Reliability section. Given a time-varying network $G = (N, A)$, where N is a finite set of nodes and A is a finite set of directed links, the time period of interest (planning horizon) is discretized into a set of small time intervals, $H = \{t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots, t_0 + T\Delta t\}$, where t_0 is the earliest possible departure time from any origin node, Δt is a small time interval during which no perceptible changes in traffic conditions and/or travel cost occur, and T is a large number such that the intervals from t_0 to $t_0 + T\Delta t$ cover the planning horizon H. The time-dependent zonal demand q^{wt} over the study horizon represents the number of individual travelers of an O-D pair $w (w \in W)$ at departure time $t (t \in T)$. A set of available modes is denoted as M. The integrated model in this study is to find a dynamic network equilibrium mode–path flow pattern by recognizing multiple dimensions of network choice behavior—that is, mode choice decision, highway user heterogeneity, and reliability of route choice.

The integrated multidimensional network choice model includes two main submodels, namely, the time-dependent mode choice stochastic user equilibrium (TDMSUE) model, and the multiclass multicriterion dynamic user equilibrium (MDUE) route choice model.

TDMSUE Model

Based on the weak law of large numbers, a mode choice probability $p_m^{wt}(y)$ can be obtained through mode flow y_m^{wt} divided by total O-D demand, q^{wt}, as follows:

$$p_m^{wt}(y) = \frac{y_m^{wt}}{q^{wt}}, \forall m \in M, w \in W, t \in T \tag{A.13}$$

The TDMSUE conditions can be stated mathematically as follows:

$$y_m^{wt} = q^{wt} \times p_m^{wt}(y), \forall m \in M, w \in W, t \in T \tag{A.14}$$

Therefore, the TDMSUE problem of interest can be formulated as the following fixed point problem. Let $\Omega(y)$ be a feasible set of mode flows.

$$\text{Find } y^* \in \Omega(y), \text{ satisfying } y^* = q \times p(y^*). \tag{A.15}$$

Solving this system of nonlinear equations (Equations A.13–A.15) will give a set of mode flows y^*, which is also the solution of the TDMSUE problem—that is, y^* would satisfy the TDMSUE condition in Equation A.14. To solve this problem by using advanced optimization-based procedures, we reformulate this problem as a gap function based nonlinear programming (NLP) in Equations A.16 through A.18, which was proposed in connection with a generalized dynamic stochastic use equilibrium problem (Zhang et al. 2008).

$$\text{Min } g_M(y) = \frac{1}{2} \times \sum_w \sum_t \sum_m \left[y_m^{wt} - q^{wt} \times p_m^{wt}(y) \right]^2 \tag{A.16}$$

subject to

$$\sum_m y_m^{wt} = q^{wt}, \forall w \in W, t \in T \tag{A.17}$$

$$y_m^{wt} \geq 0, \forall w \in W, t \in T, m \in M \tag{A.18}$$

where the objective function in Equation A.16 is a gap measure defined by summation of the square difference between the assigned mode flow, y_m^{wt}, and the expected mode flow, $q^{wt} \times p_m^{wt}(y)$, over all O-D pairs, departure times, and modes. Constraints in Equation A.17 are flow balance constraints for each O-D pair and departure time. Constraints in Equation A.18 are nonnegative mode flow constraints.

MDUE Route Choice Model

Based on the MDUE definition, the MDUE conditions can be mathematically stated as a nonlinear complementary problem (NCP) in the following:

$$\forall \alpha \in \lfloor \alpha^{\min}, \alpha^{\max} \rfloor,$$

$$x_k^{wtm}(\alpha)\left[GC_k^{wtm}(\alpha, x) - \pi^{wtm}(a, x)\right] = 0,$$
$$\forall k \in K(w, t, m), w \in W, t \in T, m \in M \tag{A.19}$$

$$GC_k^{wtm}(\alpha, x) - \pi^{wtm}(a, x) \geq 0,$$
$$\forall k \in K(w, t, m), w \in W, t \in T, m \in M \tag{A.20}$$

$$\sum_{k \in K(w,t,m)} x_k^{wtm}(\alpha) = y^{wtm}(\alpha), \forall w \in W, t \in T, m \in M \tag{A.21}$$

$$x_k^{wtm}(\alpha) \geq 0, \forall k \in K(w, t, m), w \in W, t \in T, m \in M \tag{A.22}$$

where $x = \{x_k^{wtm}(\alpha) | \forall k \in K(w, t, m), w \in W, t \in T, m \in M, \alpha \in [\alpha^{\min}, \alpha^{\max}]\}$ is a multiclass time-varying MDUE route flow vector, and $\pi^{wtm}(\alpha, x)$ is the time-varying minimum O-D generalized travel cost for each mode, evaluated at x, for the trips with the same (w, t, m, α). Constraints in Equations A.19 and A.20 are complementary constraints. Constraints in Equation A.21 are flow balance constraints. Constraints in Equation A.22 are nonnegative path flow constraints.

Given the assumptions and definition, the multiclass dynamic route choice model aims at solving the MDUE

problem, under a given road pricing scheme, to obtain a time-varying route flow vector satisfying the MDUE conditions. Based on the aforementioned NCP formulation (Equations A.19–A.22), we can derive an equivalent gap function based on NLP formulation in Equation A.23, which is an extension of an equivalent gap function–based reformulation for the dynamic user equilibrium problem (Lu et al. 2009).

$$\text{Min } g_R(x) = \sum_w \sum_t \sum_m \sum_k \int_{\alpha^{\min}}^{\alpha^{\max}} x_k^{wtm}(\alpha) \times \begin{bmatrix} GC_k^{wtm}(x,\alpha) \\ -\pi^{wtm}(x,\alpha) \end{bmatrix} d\alpha \tag{A.23}$$

subject to Equations A.20 through A.22.

Integrated Multidimensional Network Choice Model

As shown in the TDMSUE and MDUE models in the previous sections, the integrated multidimensional network choice model essentially aims to seamlessly and correctly connect the mode choice model (TDMSUE) and multimodal dynamic route choice model (MDUE). The connection between these two models is defined by the flow balance conditions as follows:

$$y_m^{wt} = \int_{\alpha^{\min}}^{\alpha^{\max}} y^{wtm}(\alpha) d\alpha = \int_{\alpha^{\min}}^{\alpha^{\max}} \left[\sum_{k \in K(w,t,m)} x_k^{wtm}(\alpha) \right] d\alpha, \quad \forall w \in W, t \in T, m \in M \tag{A.24}$$

Accordingly, the objective of the mode choice model is to obtain a TDMSUE mode flow pattern, and the objective of the multiclass dynamic route choice model is to obtain a MDUE route flow pattern with an input of the given TDMSUE mode flow pattern; in turn, the mode travel attributes (Level of Services) resulted in the MDUE route flow pattern leading to a new TDMSUE mode flow pattern. Therefore, the integrated multidimensional network choice model is mathematically formulated as follows:

$$\text{Min } g_I(y,x) = \frac{1}{2} \times \sum_w \sum_t \sum_m \left[y_m^{wt} - q^{wt} \times p_m^{wt}(y) \right]^2 \tag{A.25}$$

subject to

$$\sum_m y_m^{wt} = q^{wt}, \forall w \in W, t \in T \tag{A.26}$$

$$y_m^{wt} \geq 0, \forall w \in W, t \in T, m \in M \tag{A.27}$$

$$y_m^{wt} = \int_{\alpha^{\min}}^{\alpha^{\max}} y^{wtm}(\alpha) d\alpha = \int_{\alpha^{\min}}^{\alpha^{\max}} \left[\sum_{k \in K(w,t,m)} x_k^{wtm}(\alpha) \right] d\alpha, \quad \forall w \in W, t \in T, m \in M \tag{A.28}$$

$$x_k^{wtm}(\alpha) * \lfloor [GC_k^{wtm}(\alpha,x) - \pi^{wtm}(\alpha,x)] \rfloor = 0, \quad \forall k \in K(w,t,m), w \in W, t \in T, m \in M \tag{A.29}$$

$$GC_k^{wtm}(\alpha,x) - \pi^{wtm}(\alpha,x) \geq 0, \quad \forall k \in K(w,t,m), w \in W, t \in T, m \in M \tag{A.30}$$

$$\sum_{k \in K(w,t,m)} x_k^{wtm}(\alpha) = y^{wtm}(\alpha), \forall w \in W, t \in T, m \in M \tag{A.31}$$

$$x_k^{wtm}(\alpha) \geq 0, \forall k \in K(w,t,m), w \in W, t \in T, m \in M \tag{A.32}$$

Essentially, Equations A.28 through A.32 define *dynamic* as a multimodal route flow vector with respect to each value of time (VOT), $x^* = \{x_k^{wtm}(\alpha)^* \mid \forall k \in K(w,t,m), w \in W, t \in T, m \in M, \alpha \in [\alpha^{\min}, \alpha^{\max}]\}$, which can be obtained by solving the multimodal bicriterion dynamic use equilibrium problem.

The solution algorithm for this formulation, presented in Chapter 4, is described in greater detail in the next section of this appendix.

Solution Algorithm for the Integrated Multidimensional Network Choice Model

This section provides additional detail on the algorithmic procedures developed for the integrated multidimensional network travel choice model, with multimodal route and mode choice, introduced in Chapter 4. The algorithmic procedures were implemented for the large-scale New York Regional Best Practice Network, as described in Chapter 4.

Projected Gradient-Based Descent Direction Method to Solve the TDMSUE Problem

The TDMSUE Problem described in Equations A.16 through A.18 of this appendix can be decomposed into each O-D pair and departure time. In light of Zhang et al. (2008), we can derive a cross-set gradient of the decomposed problem as follows.

$$\begin{aligned} \nabla_{y_m^{wt}} g_{M-3}^{wt}(\mathbf{y}) &= \nabla_{y_m^{wt}} g_m^{wt}(\mathbf{y}) + \nabla_{x_{m'}^{wt}} g_{M\backslash m}^{wt}(\mathbf{y}) \\ &= [y_m^{wt} - q^{wt} \times p_m^{wt}(\mathbf{y})] \times \left[1 - q^{wt} \times \frac{\partial p_m^{wt}(\mathbf{y})}{\partial y_m^{wt}} \right] \\ &\quad + \sum_{m' \in M\backslash m} [y_{m'}^{wt} - q^{wt} \times p_{m'}^{wt}(\mathbf{y})] \times \left[-q^{wt} \times \frac{\partial p_{m'}^{wt}(\mathbf{y})}{\partial y_m^{wt}} \right] \\ &= y_m^{wt} - q^{wt} \times p_m^{wt}(\mathbf{y}) + \sum_{m' \in M} [y_{m'}^{wt} - q^{wt} \times p_{m'}^{wt}(\mathbf{y})] \\ &\quad \times \left[-q^{wt} \times \frac{\partial p_{m'}^{wt}(\mathbf{y})}{\partial y_m^{wt}} \right] \end{aligned} \tag{A.33}$$

Therefore, a projected gradient-based descent direction mode flow update scheme is as follows.

$$\mathbf{d}_y^{(n1)} = -\frac{\nabla_{\mathbf{y}} g\left(y^{(n1)}\right)}{\left\|\nabla_{\mathbf{y}} g\left(\mathbf{y}^{(n1)}\right)\right\|} \times \mathbf{y}^{(n1)} \tag{A.34}$$

$$\bar{\mathbf{d}}_y^{(n1)} = \prod_D \left(\bar{\mathbf{d}}_y^{(n1)}\right) = \left(\mathbf{I}_M - \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}\right) \times \mathbf{d}_y^{(n1)} \tag{A.35}$$

$$\mathbf{y}^{(n1+1)} = \mathbf{y}^{(n1)} + \lambda^{(n1)} \times \bar{\mathbf{d}}_y^{(n1)} \tag{A.36}$$

Multimodal Parametric Analysis Method–Based Path Generation

In actual transportation networks, the number of available routes that can carry vehicles for each O-D pair, departure time, and mode is a finite number. For a given feasible route set of (w, t, m), the least experienced generalized cost route, will be different for different travelers due to heterogeneous VOT. Following the same approach of Lu et al. (2008), we adopt the parametric shortest path approach to obtain a set of breakpoints, or values of α corresponding to the changes in the least experienced generalized cost path, that partition the feasible range of VOT α, $\lfloor[\alpha^{\min}, \alpha^{\max}]\rfloor$, into a set of subintervals, $\alpha_I = \{\alpha_0, \alpha_1, \ldots, \alpha_i, \ldots, \alpha_I | \alpha^{\min} = \alpha_0 < \alpha_1 < \cdots < \alpha_i < \cdots < \alpha_I = \alpha^{\max}\}$, and within each subinterval of VOT, $\alpha \in [\alpha_{i-1}, \alpha_i)$, $\forall i = 1, \ldots, I$, travelers are assumed to have the same path as their least experienced generalized cost path. For each α_i, O-D pair w, departure time t, and mode m, let $\tilde{K}(w, t, m, \alpha_i)$ be a restricted route set.

Integrating time-varying and heterogeneous O-D demand $y^{wtm}(\alpha)$ and route flow $x_k^{wtm}(\alpha)$ over each subinterval $[\alpha_{i-1}, \alpha_i)$, $\forall i = 1, \ldots, I$, we obtain the time-varying demand vector $y^{wtm}(\alpha_I) = \{y^{wtm}(\alpha_i) | \forall i = 1, \ldots, I\}$, and route flow vector, $x_k^{wtm}(\alpha_I) = \{x_k^{wtm}(\alpha_i) | \forall i = 1, \ldots, I\}$ for each user class i, as in Equations A.37 and A.38, respectively.

$$\begin{aligned} y^{wtm}(\alpha_i) &= \int_{\alpha_{i-1}}^{\alpha_i} y^{wtm}(\alpha) d\alpha = \int_{\alpha_{i-1}}^{\alpha_i} y^{wtm} \times \varphi(\alpha) d\alpha \\ &= y^{wtm} \times [\Phi(\alpha_i) - \Phi(\alpha_{i-1})] \end{aligned} \tag{A.37}$$

$$\begin{aligned} x_k^{wtm}(\alpha_i) &= \int_{\alpha_{i-1}}^{\alpha_i} x_k^{wtm}(\alpha) d\alpha = \int_{\alpha_{i-1}}^{\alpha_i} x_k^{wtm} \times \varphi(\alpha) d\alpha \\ &= x_k^{wtm} \times [\Phi(\alpha_i) - \Phi(\alpha_{i-1})] \end{aligned} \tag{A.38}$$

Equations A.37 and A.38 can be rewritten as Equations A.39 and A.40, respectively.

$$y^{wtm} = \sum_{i=1}^{I} y^{wtm}(\alpha_i), \forall w \in W, t \in T, m \in M \tag{A.39}$$

$$x_k^{wtm} = \sum_{i=1}^{I} x_k^{wtm}(\alpha_i), \forall k \in \tilde{K}(w, t, m), w \in W, t \in T, m \in M \tag{A.40}$$

Restricted MDUE Problem

In light of the multimodal parametric analysis method and the integrated multidimensional network choice model, we can derive a restricted multiclass dynamic use equilibrium (RMDUE) problem as follows in Equation A.41:

$$\begin{aligned} \text{Min } g_{M-4}(x, \alpha_I) = &\sum_w \sum_t \sum_m \sum_k \sum_i x_k^{wtm}(\alpha_i) \\ &\times \left[GC_k^{wtm}(x, \alpha_i) - \pi^{wtm}(x, \alpha_i)\right] \end{aligned} \tag{A.41}$$

subject to

$$\begin{aligned} &GC_k^{wtm}(x, \alpha_i) - \pi^{wtm}(x, \alpha_i) \geq 0, \\ &\quad \forall k \in \tilde{K}(w, t, m, \alpha_i), w \in W, t \in T, m \in M, i \in I \end{aligned} \tag{A.42}$$

$$\sum_{k \in \tilde{K}(w,t,m,\alpha_i)} x_k^{wtm}(\alpha_i) = y^{wtm}(\alpha_i), \forall w \in W, t \in T, m \in M, i \in I \tag{A.43}$$

$$x_k^{wtm}(\alpha_i) \geq 0, \forall k \in \tilde{K}(w, t, m, \alpha_i), w \in W, t \in T, m \in M, i \in I \tag{A.44}$$

As such, we can transform the continuous complementarity NLP problem into a discrete problem, which can be solved in a column generation solution framework.

Projected Gradient-Based Descent Direction Method to Solve the RMDUE Problem

The RMDUE problem can be decomposed into each O-D pair and departure time. Following Lu et al. (2008), we can derive a projected gradient-based descent direction method to solve the decomposed problem as follows in Equations A.45 and A.46:

$$d_x^{(n2)}(\alpha_i) = -1 \times \frac{GC_k^{wtm}(x, \alpha_i) - \pi^{wtm}(x, \alpha_i)}{GC_k^{wtm}(x, \alpha_i)} \tag{A.45}$$

$$x_k^{wtm^{(n2+1)}}(\alpha_i) = \begin{cases} x_k^{wtm^{(n2)}}(\alpha_i) + \lambda^{(n2)} \times d_x^{(n2)}(\alpha_i), \\ \qquad \forall k \in K(w, t, m, \alpha_i) \setminus K_\pi(w, t, m, \alpha_i) \\ x_k^{wtm^{(n2)}}(\alpha_i) + \dfrac{\sum_{k' \in K(w,t,m,\alpha_i) \setminus K_\pi(w,t,m,\alpha_i)} x_{k'}^{wtm^{(n2+1)}}(\alpha_i)}{|K_\pi(w, t, m, \alpha_i)|}, \\ \qquad \forall k \in K_\pi(w, t, m, \alpha_i) \end{cases} \tag{A.46}$$

Simulation-Based Iterative Solution Framework

The TDMSUE–MDUE problem consists of finding both equilibrium travelers' mode choice and equilibrium vehicles' route choice with a given time-dependent traveler demand. The TDMSUE problem is solved by a projected gradient-based descent direction method. However, it is difficult to enumerate the complete set of feasible routes for solving the MDUE problem in a practical size transportation network. To capture the individual choice behavior and traffic dynamics, the simulation-based dynamic traffic assignment (DTA) algorithmic framework disaggregates the O-D demands into individual vehicles, and only a portion of paths would have a nonzero probability to carry vehicles in an MDUE solution. This study uses the trajectories of vehicles as a proxy to store the feasible path set, namely, the vehicle-based implementation technique, to save computer memory and eliminate some unrealistic paths. To avoid explicit enumeration of all feasible routes, this study applies a column generation approach to generate a representative subset of paths with competitive cost and augment the feasible path set. The parametric analysis method (PAM) is applied to obtain a set of breakpoints that partition the entire VOT interval into multiple subintervals. A projected descent direction method is used to solve the resulting MDUE problem in a restricted/reduced path set, called an RMDUE problem, in Equations A.41 through A.44.

The simulation-based column generation iterative solution framework for the TDMSUE–MDUE problem includes four main steps: (1) input and initialization, (2) nested logit mode choice, (3) ride-sharing choice and vehicle generation, and (4) multidimensional simulation-based dynamic micro-assignment. This solution algorithm is shown in Figure A.1.

Application to New York Regional Network: Calibration of Time-Dependent O-D Demand with Multiple Vehicle Types

This section describes the steps and procedures involved in developing the application of the integrated user choice processes and network modeling procedures presented in Chapter 4 to the New York region "best practice" network. As noted, this is the largest actual network application of simulation-based network equilibrium procedures reported to date. Applications of this scale and magnitude pose additional challenges beyond the usual steps encountered in applying dynamic modeling procedures using data initially developed for static model application. These are documented in this section, with particular emphasis on the innovations developed to estimate unknown time-dependent demand patterns for the DTA model.

Building a Large-Scale Network Model: Summary of Challenges

In general, because of their ability to represent network operational characteristics, simulation-based dynamic traffic assignment models require more in-depth network information than comparable static assignment models. Traffic control signs and signals, left turns, and other movement capabilities at a node are mainly (only crudely) represented in the link performance function in a static network, whereas a dynamic model requires more accurate information on junction control and allowed movements at each phase at a signalized intersection, as well as careful definition of each downstream movement at a node.

Basically, there is no direct method of (correctly) converting a static network model into a dynamic network model in a single attempt by only using the existing data obtained from the provided static network model database. Smart conversion of existing database, use of external information sources, and more importantly, use of engineering judgment, are essential elements of building a large-scale dynamic network model.

In summary, models developed for static assignment application generally lack several essential elements for dynamic network analysis, including

- Oversimplified representation of junctions, especially freeway interchanges for correct operational simulation;
- Absence or incorrect control information at junctions, and lack of reliable electronic database of control devices and control parameters at signalized junctions;
- Definition of origin and destination zones, including treatment/connection of centroids and external traffic generators;
- Insufficient specification of the operational attributes of links and junctions for the purpose of traffic simulation; and
- Absence of time-varying O-D information, which must be synthesized from available static matrices, coupled with traffic counts sometimes taken in mutually different time periods.

Conversion of Existing Network for Dynamic Analysis

The regional New York best practice model (NYBPM) includes 28 counties from three different states divided into 3,586 internal traffic analysis zones (TAZs):

- 10 counties from the New York Metropolitan Transportation Council (NYMTC) area;

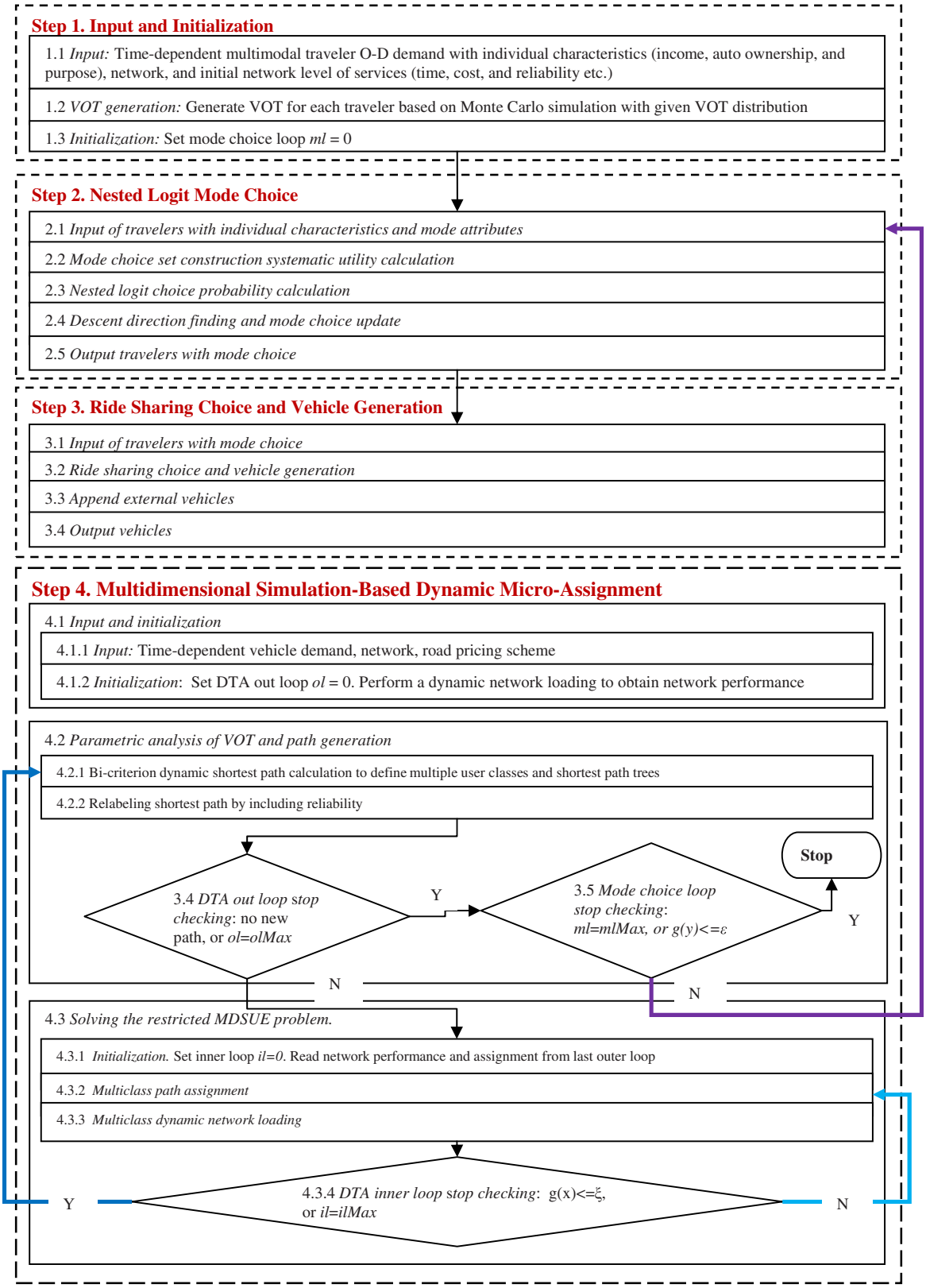

Figure A.1. Simulation-based column-generation solution framework.

- 2 other counties from New York State;
- 13 counties from the North Jersey Transportation Planning Authority (NJTPA) area;
- 1 other county from the state of New Jersey; and
- 2 counties from the state of Connecticut.

The zones are mainly concentrated in New York City (NYC). See Figure A.2 for a general view of the NYBPM network modeled in TransCAD.

- 2,449 zones from New York State;
- 740 zones from the state of New Jersey;
- 397 zones from the state of Connecticut; and
- 111 external zones for travel entries to and exits from the network.

Furthermore, the NYBPM network consists of

- 53,395 links; and
- 31,812 nodes.

The DTA model converted from the static TransCAD model can be seen in Figures A.2 and A.3.

The existing network was made available in TransCAD but is limited for static assignment application, and hence suffers from the limitations just noted. These had to be overcome through a systematic process of verifying and correcting the topology and connectivity of the network, assigning correct operational attributes for simulation purposes, determining appropriate junction control (often with visual verification through aerial photography, using sources such as Google Earth), defaulting signal control parameters, and estimating time-dependent O-D patterns using a state of the art methodology developed for this purpose.

In the existing static model, most arterial-freeway interchanges are designed properly in NYC, whereas the interchanges in New Jersey and Connecticut are designed as at-grade intersections. In a dynamic model, if an arterial and a freeway link, or two freeway links directly meet at a node, then the movements will be incorrect; for example, a left-turning vehicle would be able to block the opposing traffic, which is impossible on a full-access controlled road. This would result in an unrealistic dynamic traffic assignment. Therefore, conversion of these intersections into interchanges was necessary and has been done manually using the "Create Interchanges" toolbox of TransCAD. A before–after example of such a case is shown in Figure A.4.

Table A.1 lists the required and optional input data files for DYNASMART-P. Figure A.5 depicts the flowchart of the conversion methodology.

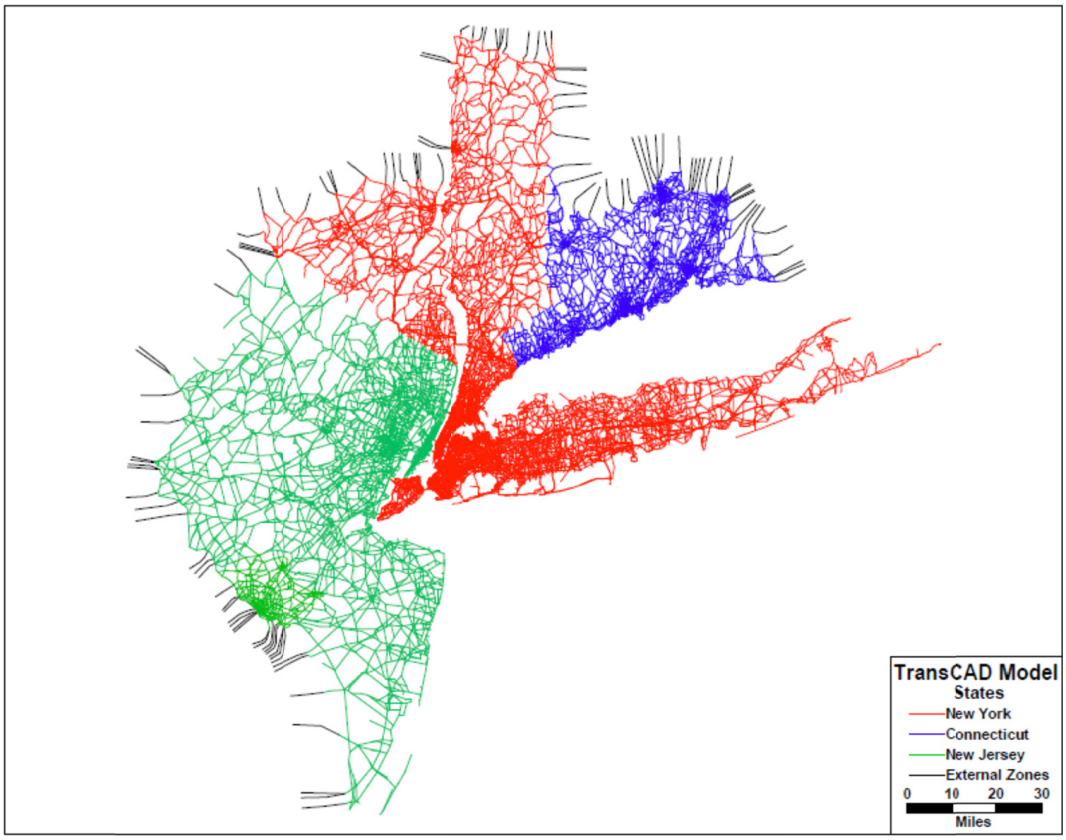

Figure A.2. TransCAD model of the NYBPM network.

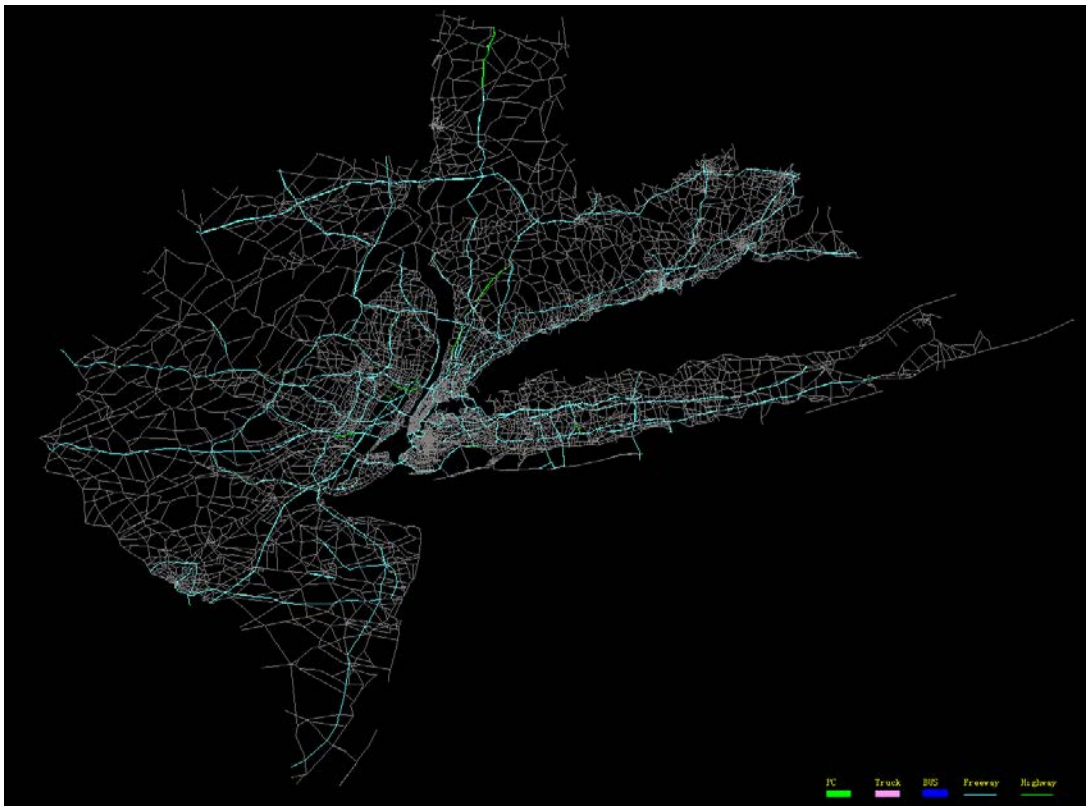

Figure A.3. DYNASMART-P model of the New York network.

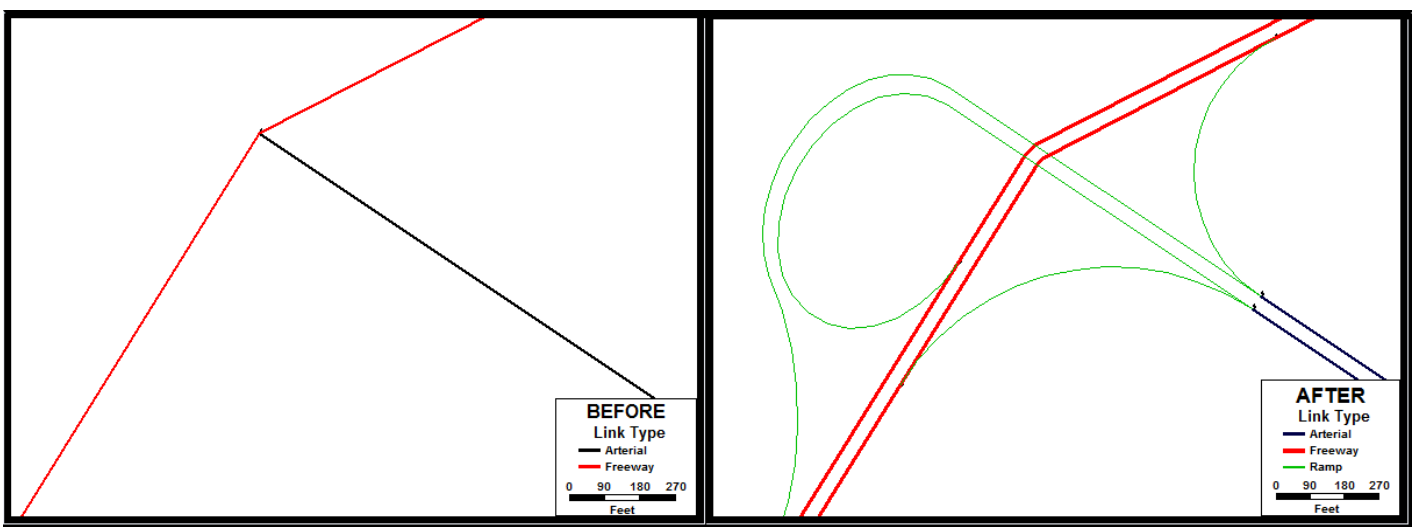

Figure A.4. Ramp correction for realistic turn movements.

Table A.1. Input Files for DYNASMART-P

Input File	Description	Status
Network Data		
xy.dat	Contains the coordinates of the physical nodes to be used by the DYNASMART-P GUI	Optional
linkxy.dat	Contains links' starting and ending nodes. For more accurate representation, users can specify as many feature points as needed to create curvatures	Optional
linkname.dat	Contains names of links (i.e., street names)	Optional
network.dat	Contains information regarding the network configuration, zoning, node, and link characteristics	Required
movement.dat	Contains information regarding the allowed movements of vehicles (right turns, left turns, through, and others)	Required
TrafficFlowModel.dat	Contains information regarding the type of speed-density models specified and their corresponding parameters	Required
GradeLengthPCE.dat	Contains information regarding the effect of grade length and truck percentage on passenger car equivalent (PCE) factors	Required
Control Data		
control.dat	Contains information regarding the type of traffic control at each node; phasing information, if the intersection is signalized; major and minor approaches if the intersection has yield or two-way stop sign	Required
leftcap.dat	Contains information regarding the left-turn capacity at signalized intersections	Required
StopCap2Way.dat	Contains information regarding the capacity of approaches at two-way stop-controlled intersections	Required
StopCap4Way.dat	Contains information regarding the capacity of approaches at all-way stop-controlled intersections	Required
YieldCap.dat	Contains information regarding the capacity of approaches at yield-controlled intersections	Required
Traffic Management Data		
scenario.dat	Contains information regarding the basic simulation variables	Required
vms.dat	Contains information regarding the location, type, and time of activation for variable message signs	May be empty if no signs
WorkZone.dat	Contains information regarding the number of work zone links, duration, capacity reduction, new speed limit, and discharge rate	May be empty
incident.dat	Contains information about capacity reduction (fraction) and duration of incident on links	May be empty
pricing.dat	Contains information regarding the pricing on the regular and high-occupancy toll (HOT)/high-occupancy vehicle (HOV) lanes	May be empty
ramp.dat	Contains information regarding ramp locations, detector locations, metering rate, and its timing	May be empty
bus.dat	Contains information regarding the buses, including the trajectories, location of stops, dwell time	May be empty
output_option.dat	Contains information regarding the frequency of writing to output files	Required
system.dat	Contains information regarding selection of the solution mode, length of the planning horizon, aggregation interval, and assignment interval	Required
Demand Data		
zone.dat	Contains information that describes the zonal boundaries	Optional
demand.dat	Contains information regarding the temporal and spatial distribution of vehicular demand	Required
demand_HOV.dat	Contains information regarding the temporal and spatial distribution of HOV demand	Required
demand_truck.dat	Contains information regarding the temporal and spatial distribution of truck demand	Required
destination.dat	Contains information regarding destinations in the network	Required
origin.dat	Contains information regarding the generation links in the network	Required
vehicle.dat	Contains information regarding the number of vehicles to be loaded	May be empty
path.dat	Contains the vehicle trajectory to be used in conjunction with vehicle.dat	May be empty
SuperZone.dat	Contains the mapping of zones to super zones	Optional

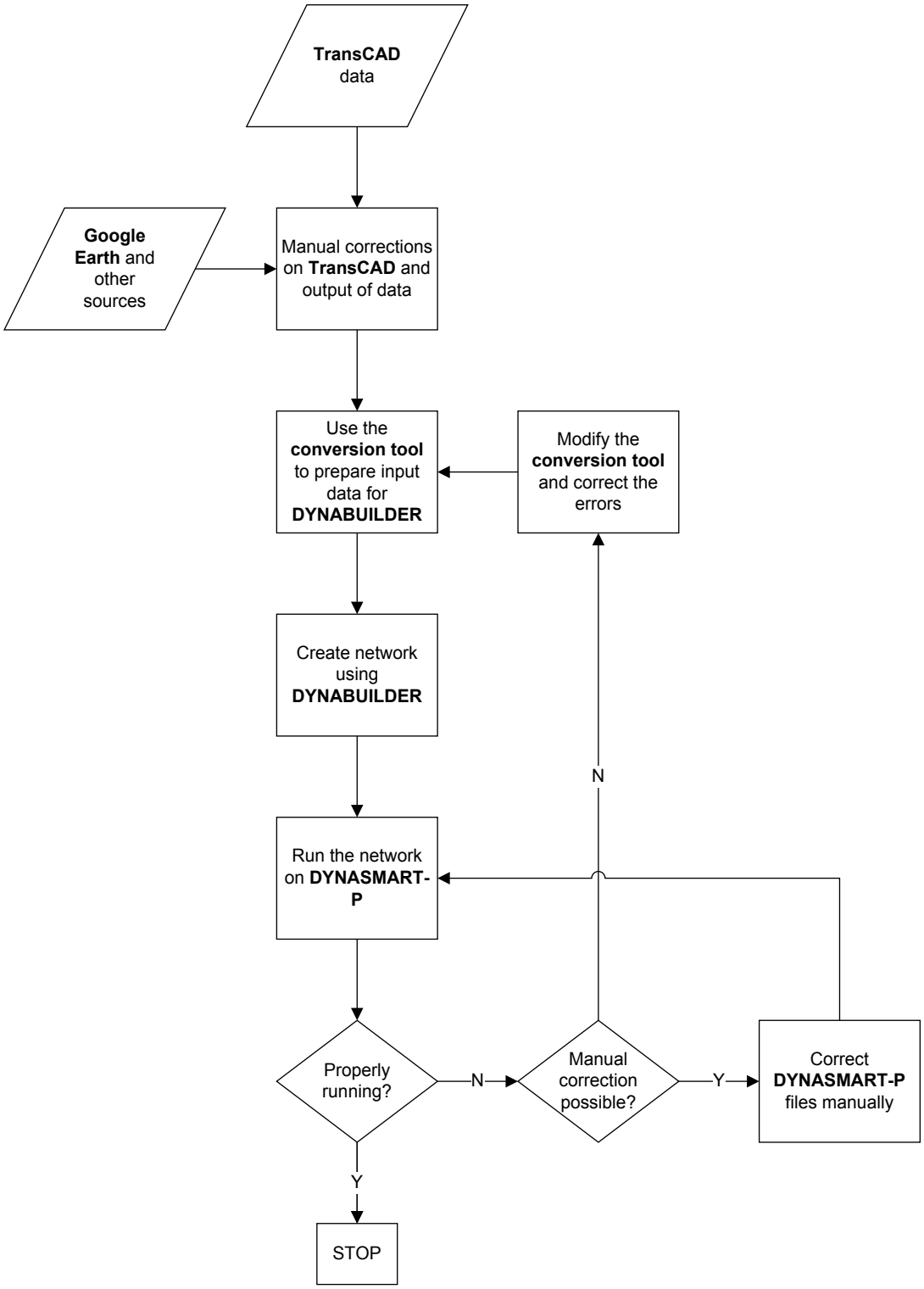

Figure A.5. Flowchart for the conversion from the static to the dynamic network model.

Preparation of DYNASMART-P data files for the simulation and graphical user interface (GUI) consists of several steps. The main tool for this conversion is a software package called DYNABUILDER, which is capable of converting many networks from different platforms into a DYNASMART-P network. However, DYNABUILDER requires input files in a certain format. The conversion of the TransCAD dataset into DYNABUILDER input was established using several different codes and macros based on the need. The package of these custom-made codes and macros are called "conversion tool" in the flowchart.

Signal information was provided as a link feature, which is actually a node feature. Using the conversion tool, the downstream nodes of the links containing signals were assigned traffic signals, ensuring that the node has more than one downstream movement and no freeway link is involved. Arterials with more than three lanes are also assigned signals at their downstream nodes using the same conditions. The phasing and movements at the signalized intersections are done by DYNABUILDER. Because no further information is available for the vast number of signalized intersections, all of them are assigned to be actuated signals with default minimum and maximum green times and amber times.

Two-way stop and yield signs are assigned also by the conversion tool according to intersection configuration. Ramps entering highways are assigned yield signs, and arterial intersections with different lane numbers on the upstream approaches are assigned two-way stop signs. The major/minor approach assignment is also done by the conversion tool.

Geometric configuration of the network and movements at the nodes are done by DYNABUILDER using the provided input by the conversion tool. Another important point is the removal of internal and external centroid nodes and internal and external centroid connector links. DYNASMART-P only requires information for physical nodes and links. It creates centroids and connectors implicitly for each zone. Furthermore, all bidirectional links were dualized—that is, converted into two one-directional links, which is another requirement for DYNASMART-P.

As a result, the DYNASMART-P network has the following properties.

Zone Information

- 3,697 zones
  - 3,586 internal; and
  - 111 external.

Node and Control Information

- 28,406 nodes
  - 3,816 uncontrolled;
  - 2,625 yield signed;
  - 12,944 all-way stop signed;
  - 8,054 actuated controlled; and
  - 967 two-way stop signed.

Link and Type Information

- 68,490 links
  - 6,026 freeways;
  - 169 freeway HOV links;
  - 56,102 arterials;
  - 37 HOV arterial links;
  - 150 highways;
  - 2,688 on-ramps; and
  - 3,318 off-ramps.

Pricing Information

There are 297 tolled links:

- 291 static tolling; and
- 6 dynamic tolling.

As seen in Figure A.6, most of the pricing is nondistance based, except along the I-95 New Jersey Turnpike corridor.

- The George Washington Bridge, Lincoln Tunnel, Holland Tunnel, Goethals Bridge, Outerbridge Crossing, and Bayonne Bridge are dynamically tolled bridges.
- The Verrazano-Narrows Bridge, Bronx–Whitestone Bridge, Brooklyn–Battery Tunnel, Queens Midtown Tunnel, Throgs Neck Bridge, Triborough Bridge, Marine Parkway-Gil Hodges Memorial Bridge, Cross Bay Veterans Memorial Bridge, and Henry Hudson Bridge are the bridges and tunnels tolled in New York Metropolitan Area.
- The Tappan Zee Bridge, Bear Mountain Bridge, Kingston Rhinecliff Bridge, Mid Hudson Bridge, and Newburgh Beacon Bridge are the tolled bridges in New York State.

Methodology for Calibration of O-D Demand for Dynamic Analysis

Given static O-D demand information and time-dependent link measurements, the dynamic O-D demand estimation procedure aims to find a consistent time-dependent O-D demand table that minimizes (1) the deviation between estimated link flows and observed link counts, and (2) the deviation between the estimated demand and the target demand (based on the static demand matrix). The induced network flow pattern can be expressed in terms of path flows and link flows.

In a dynamic context, especially in congested networks, elements of the mapping matrix between O-D demand and link flows are not constant and are, themselves, a function of the unknown O-D demand values. A bi-level dynamic O-D estimation formulation is adapted here to integrate the

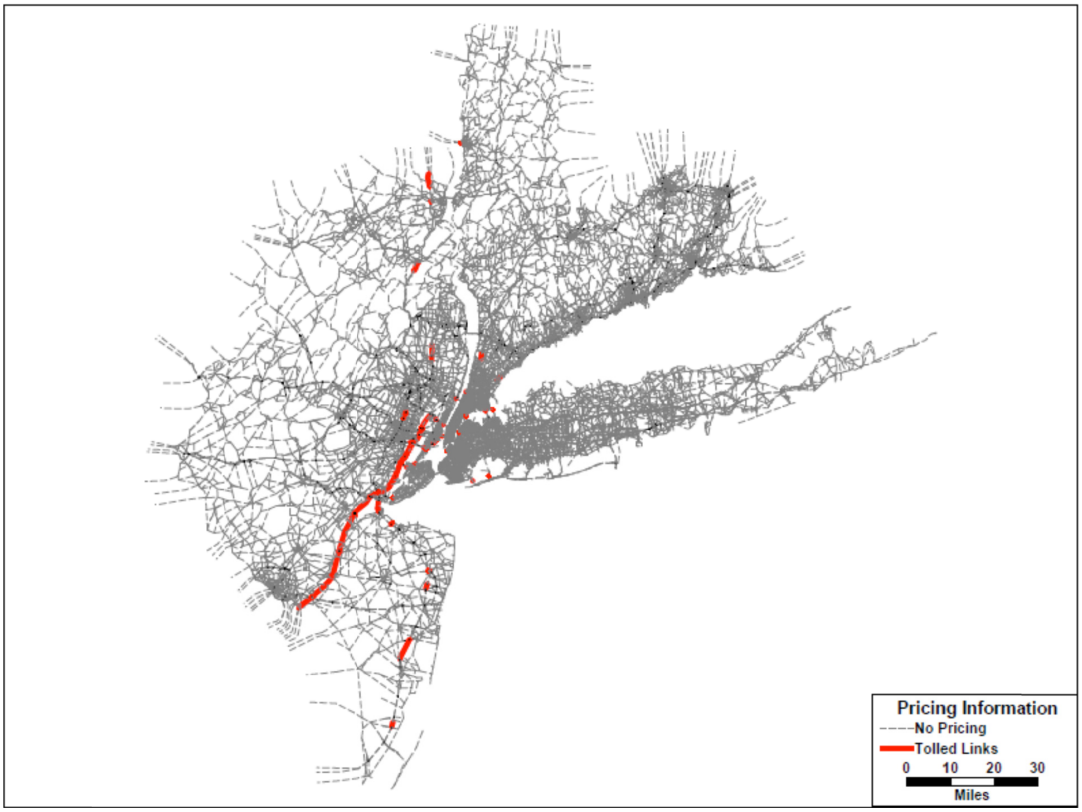

Figure A.6. Pricing information.

dynamic traffic assignment constraint. Specifically, the upper-level problem seeks to estimate the dynamic O-D trip desires based on given link counts and flow proportions, subject to nonnegativity constraints for demand variables. The flow proportions are, in turn, generated from the dynamic traffic network loading problem at the lower level, which is solved by a DTA simulation program (DYNASMART-P), with a dynamic O-D trip table calculated from the upper level. The weights $w_1 = (1 - w)$ and $w_2 = w$ in the combined deviations could be interpreted as the decision maker's relative preference or importance belief for the different objectives; they could also be considered as the dispersion scales for the first and second error terms in the ordinary least-squares estimation procedure.

Upper Level: Constrained Ordinary Least-Squares Problem

$$w_1 \sum_{l,t} \left[\sum_{i,j,\tau} p_{l,t,i,j,\tau} \times q_{i,j,\tau} - c_{l,t} \right]^2 + w_2 \sum_{i,j} \left[\sum_{\tau} q_{i,j,\tau} - d_{i,j} \right]^2$$

Subject to nonnegativity constraints $q_{i,j,\tau} \geq 0 \quad \forall i, j, \tau$

where

$q_{i,j,\tau}$ = estimated traffic demand from zone i to zone j at departure interval τ;

$c_{l,t}$ = measured traffic flows on link l at observation interval t;

$d_{i,j}$ = induced traffic demand from zone i to zone j (target demand);

$p_{l,t,i,j,\tau}$ = time-dependent link-flow proportions—that is, fraction of vehicular demand flows from origin i to destination j, starting their trips during departure interval τ, contributing to the flow on link l during observation interval t; and

w_1, w_2 = weighting factors for the first and second objective functions in the weighted objective function, where w_1 is weight for the deviations from observed link flows and the w_2 is the deviations from target demand.

Lower Level: Dynamic Traffic Assignment Problem

$$[P] = DTA(Q)$$

where

P = link proportion matrices contain link-flow proportions $p_{l,t,i,j,\tau}$;

Q = time-dependent O-D demand vector contains elements $[q_{i,j,\tau}]$ for the subarea network; and

DTA = function of dynamic traffic assignment process.

Three types of input are required for this task, namely the link proportions, link observations, and an initial estimate of the O-D demand matrix (target matrix). DYNASMART-P is first run with the target matrix. Then the vehicle trajectory file (DYNASMART-P output) is post-processed to determine the link proportions. The O-D estimation module is then executed and a more-consistent O-D demand matrix is obtained and fed back into the system until convergence. The general procedure for this task is depicted in Figure A.7.

Calibration Results and Validation Tests

Much effort was spent to come up with the best representative O-D demand matrix. Recognizing some of the data limitations described earlier, it was still possible to develop and calibrate a reliable dynamic traffic assignment tool that represents the dynamics of traffic in the study area to a reasonable degree, and allows meaningful comparative analysis of alternative scenarios.

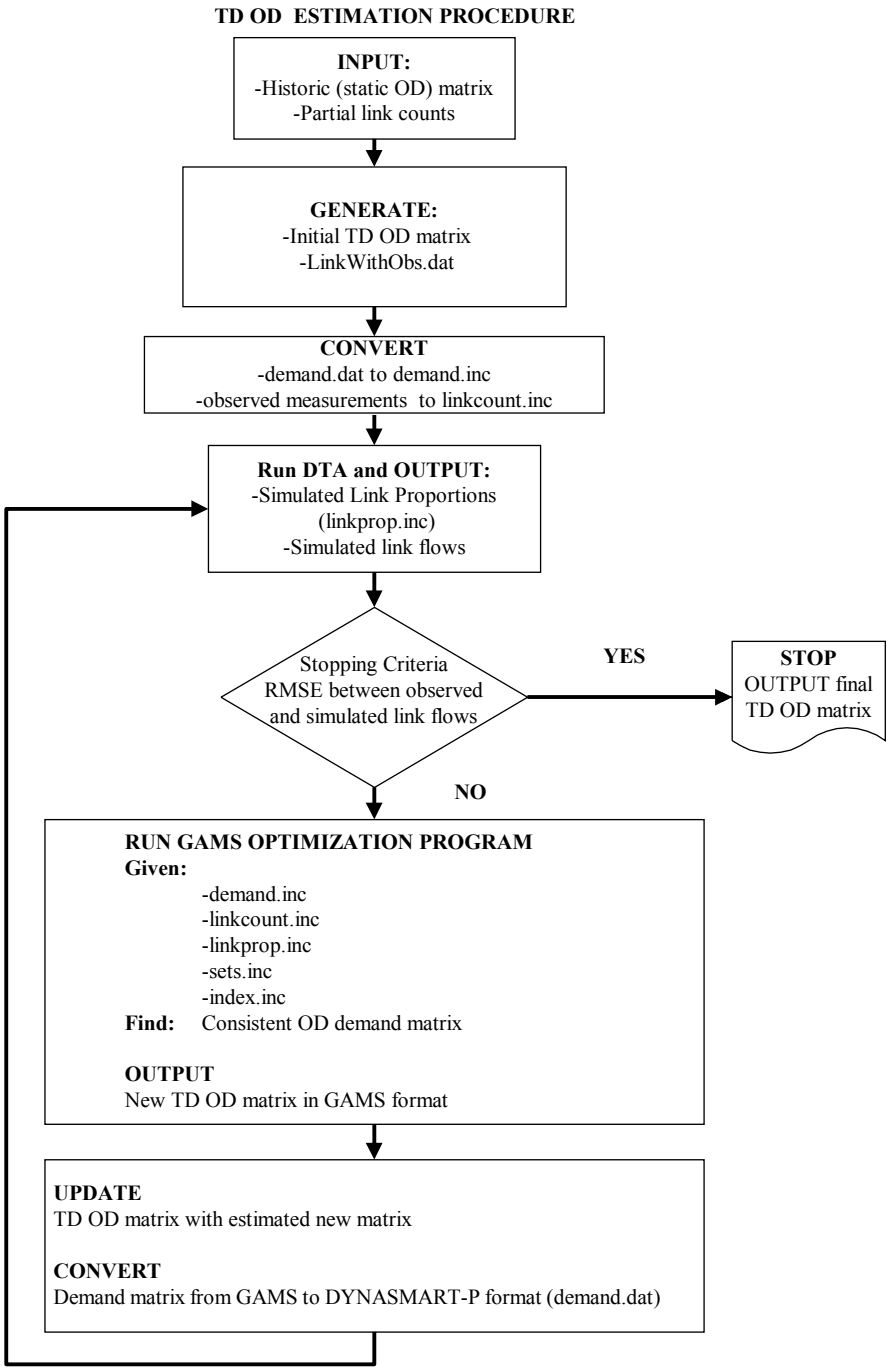

Figure A.7. Schematic of the procedure for O-D demand estimation.

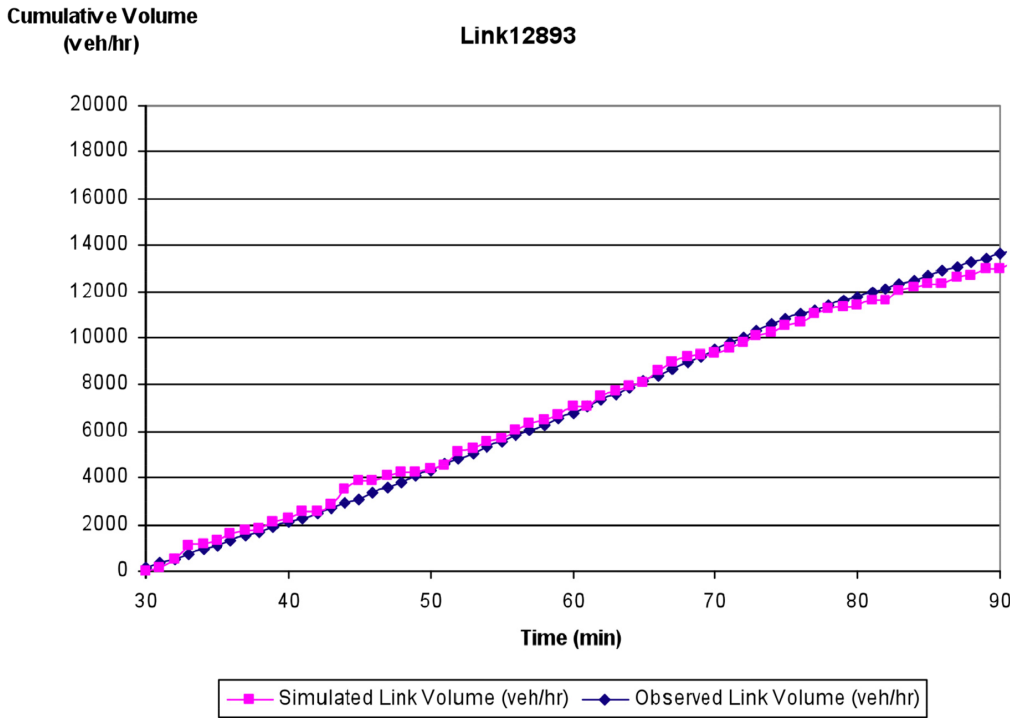

Figure A.8. Sample of simulated link volumes versus observed link volumes.

To evaluate the performance of the procedure, the root mean squared error between observed link volumes and simulated link volumes is used as an overall measure of effectiveness.

Validation against individual link counts was performed for selected links. Cumulative curves provide insight into the ability of the resulting assignment to capture the link flow volumes. The results are satisfactory in light of the available data, from the aggregate initial demand matrix to the link counts used, and provide encouraging indications for the ability of the DTA tool to support the intended analysis of traffic patterns under various scenarios. For an example of the results of the simulated link volumes compared with observed link volumes, see Figure A.8.

References

Abdelghany, A. F., and H. S. Mahmassani. 2003. Temporal-Spatial Micro-assignment and Sequencing of Travel Demand with Activity-Trip Chains. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1831,* Transportation Research Board of the National Academies, pp. 89–97.

Abou-Zeid, M., T. F. Rossi, and B. Gardner. 2006. Modeling Time-of-Day Choice in Context of Tour- and Activity-Based Models. *Transportation Research Record: Journal of the Transportation Research Board, No. 1981,* Transportation Research Board of the National Academies, Washington, D.C., pp. 42–49.

Abrantes, P., and M. Wardman. 2011. Meta-analysis of UK values of travel time: An update. *Transportation Research Part A: Policy and Practice,* Vol. 45, No. 1, Elsevier.

Adler, T., Y. Dehghani, M. Doherty, J. Klodzinski, and W. Olsen. 2005. A Multi-Period Toll Mode Choice Model for Florida's Turnpike Enterprise. *Proc., 10th TRB Transportation Planning Applications Conference,* Portland, Ore.

Adler, T., W. Olsen, and Y. Dehghani. 2004. A Facility-Level Model of Peak Spreading. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Appiah, J. 2004. *An Examination of Factors Affecting High Occupancy/Toll Lane Demand.* MS thesis. Texas A&M University, College Station.

Ariely, D. 2010. *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions.* Harper Collins.

Arup. 2003. *Frameworks for Modeling the Variability of Journey Times on the Highway Network.* London.

Asakura, Y. 1998. Reliability Measures of an Origin and Destination Pair in a Deteriorated Road Network with Variable Flows. In *Transportation Networks: Recent Methodological Advances* (M. G. N. Bell, ed.), Pergamon Press, Oxford, pp. 273–287.

Axhausen, K., S. Hess, A. Konig, G. Abay, J. Bates, and M. Bierlaire. 2007. State-of-the-Art Estimates of Swiss Value of Travel Time Savings. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Balke, K., P. Songchitruksa, N. Chaudhary, C.-L. Chu, S. Sunkari, P. Nelson, S. Kuchangi, V. Tyagi, and D. Swaroop. 2007. *Development of a Prototype Dynamic Congestion and Incident Prediction System.* Report No. FHWA/TX-07/0-4946-2. Texas Transportation Institute, College Station.

Banks, J. 1991. Two-Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering? In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., pp. 83–90.

Bates, J. 2000. Reliability: The Missing Model Variable. Presented at 9th International Conference on Travel Behavior, Gold Coast, Queensland, Australia, 2000.

Bates, J., I. Black, J. Fearon, C. Gilliam, and S. Porter. 2002. Supply Models for Use in Modeling the Variability of Journey Times on the Highway Network. *Proc., European Transport Conference,* Association for European Transport, Cambridge, United Kingdom.

Bates, J., J. Polak, P. Jones, and A. Cook. 2001. The Valuation of Reliability for Personal Travel. *Transportation Research Part E,* Vol. 37, No. 2–3, pp. 191–229.

Batley, R. 2007. Marginal Valuations of Travel Time and Scheduling, and the Reliability Premium. *Transportation Research Part E,* Vol. 43, No. 4, pp. 387–408.

Batley, R., and A. Daly. 2004. Departure Time Choice Under Uncertainty: A Model of Ransom Utility Maximization for Wagers. Presented at European Transport Conference, Strasbourg, France.

Batley, R., S. Grant-Muller, J. Nellthorp, G. de Jong, D. Watling, J. Bates, S. Hess, and J. Polak. 2008. *Multimodal Travel Time Variability: Final Report.* Department for Transport, London.

Bekhor, S., D. Christoph, and K. W. Axhausen. 2011. Integration of Activity-Based with Agent-Based Models: Case of Tel Aviv, Israel. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2255,* Transportation Research Board of the National Academies, Washington, D.C., pp. 38–47.

Bell, M. G. H., and C. Cassir. 2000. *Reliability of Transport Networks.* Research Studies Press, Ltd., Baldock, United Kingdom.

Ben-Akiva, M., and M. Abou-Zeid. 2007. Methodological Issues in Modeling Time-of-Travel Preferences. *11th World Conference on Transport Research,* Berkeley, Calif.

Black, I. G., and J. G. Towriss. 1993. *Demand Effects of Travel Time Reliability.* Centre for Logistics and Transportation, Cranfield Institute of Technology, Cranfield, United Kingdom.

Bogers, E. A. I., H. W. C. van Lint, and H. J. van Zuylen. 2008. Reliability of Travel Time: Effective Measures from Behavioral Point of View. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2082,* Transportation Research Board of the National Academies, Washington, D.C., pp. 27–34.

Borjesson, M., M. Fosgerau, and S. Algers. 2012. On the Income Elasticity of the Value of Travel Time. *Transportation Research Part A: Policy and Practice,* Vol. 46, No. 2, Elsevier, pp. 368–377.

Boxill, S. A., and L. Yu. 2000. *An Evaluation of Traffic Simulation Models for Supporting ITS Development.* Report No. SWUTC/00/167602-1. Center for Transportation Training and Research, Texas Southern University, Houston.

Boyce, D., C. R. O'Neil, and W. Scherr. 2007. New Computational Results on Solving the Sequential Procedure with Feedback. Presented at 11th TRB National Planning Applications Conference, Daytona Beach, Fla.

Boyles, S., S. Ukkusuri, S. T. Waller, and K. Kockelman. 2008. A Comparison of Static and Dynamic Traffic Assignment Under Tolls in the Dallas-Fort Worth Region. *Transportation Research Board Conference Proceedings,* Vol. 2, No. 42, pp. 114–117.

Boyles, S., S. V. Ukkusuri, S. T. Waller, and K. Kockelman. 2006a. Examining the Benefit of Accounting for Traffic Dynamics in Congestion Pricing Applications. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C.

Bradley, M. A., and H. F. Gunn. 1990. *Stated Preference Analysis of Values of Travel Time in the Netherlands.* Transportation Research Record No. 1285, Transportation Research Board of the National Academies, Washington, D.C., pp. 78–88.

Brennand, A. W. 2011. *Incorporating Travel Time Reliability in the Estimation of Assignment Models.* New Zealand Transport Agency Research Report 464. Wellington, New Zealand.

Brilon, W., J. Geistefeldt, and M. Regler. 2005. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. In *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction* (H. S. Mahmassani, ed.), Elsevier.

Brownstone, D., A. Chosh, T. F. Golob, C. Kazimi, and D. Van Amelsfort. 2003. Drivers' Willingness-to-Pay to Reduce Travel Time: Evidence from the San-Diego I-15 Congestion Pricing Project. *Transportation Research Part A,* Vol. 37, No. 4, pp. 373–388.

Brownstone, D., and K. A. Small. 2005. Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A,* Vol. 39, No. 4, pp. 279–293.

Burghout, W. 2004. *Hybrid Microscopic–Mesoscopic Traffic Simulation.* PhD dissertation. Royal Institute of Technology, Stockholm, Sweden.

Burris, M., and J. Appiah. 2004. An Examination of Houston's QuickRide Participants by Frequency of QuickRide Usage. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1864,* Transportation Research Board of the National Academies, Washington, D.C., pp. 22–30.

Burris, M., K. Konduru, and C. Swenson. 2004. Long-Run Changes in Driver Behavior Due to Variable Tolls. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1864,* Transportation Research Board of the National Academies, Washington, D.C., pp. 78–85.

Burris, M. W., and R. M. Pendyala. 2002. Discrete Choice Models of Traveler Participation in Differential Time of Day Pricing Programs. *Transport Policy,* Vol. 9, No. 3, pp. 241–251.

Cambridge Systematics, Inc., Mark Bradley Research and Consulting, ECO Northwest, and MORPACE International. 2007. *PSRC 2006 Household Activity Survey Analysis Report.* Puget Sound Regional Council and Washington State Department of Transportation, Olympia, WA.

Cambridge Systematics, Inc., Dowling Associates, Inc., System Metrics Group, Inc., and Texas Transportation Institute. 2007. *NCHRP Report 618: Cost-Effective Performance Measures for Travel Time Delay, Variation, and Reliability.* Transportation Research Board of the National Academies, Washington, D.C.

Cambridge Systematics, Inc., and Texas Transportation Institute. 2005. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation: Final Report.* Federal Highway Administration, U.S. Department of Transportation.

Cassidy, M., and R. Bertini. 1999. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B,* Vol. 33, No. 1, pp. 25–42.

Castiglione, J., B. Grady, M. Outwater, and S. Lawe. 2012. Sensitivity Testing and Assessment of an Integrated Regional Travel Demand and Traffic Microsimulation Model. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C.

Chan, K. S., and W. H. K. Lam. 2005. Impact of Road Pricing on the Network Reliability. *Journal of the Eastern Asia Society for Transportation Studies,* Vol. 6, pp. 2060–2075.

Chang, G.-L., H. S. Mahmassani. 1988. Travel Time Prediction and Departure Time Adjustment Behavior Dynamics in a Congested Traffic System. *Transportation Research Part B: Methodological,* Vol. 22, No. 3, Elsevier, pp. 217–232.

Chang, G.-L., H. S. Mahmassani, and R. Herman. 1985. A Macroparticle Traffic Simulation Model to Investigate Peak-Period Commuter Decision Dynamics. In *Transportation Research Record 1005,* TRB, National Research Council, Washington, D.C., pp. 107–120.

Chen, A., Z. Ji, and W. Recker. 2003. Effect of Route Choice Models on Estimation of Travel Time Reliability Under Demand and Supply Variations. In *Network Reliability of Transport* (M. G. H. Bell and Y. Iida, eds.), Kyoto, Japan, pp. 93–118.

Chen, A., H. Yang, H. K. Lo, and W. H. Tang. 1999. A Capacity-Related Reliability for Transportation Networks. *Journal of Advanced Transportation,* Vol. 33, No. 2, pp. 183–200.

Chen, A., H. Yang, H. K. Lo, and W. H. Tang. 2002. Capacity Reliability for a Road Network: An Assessment Methodology and Numerical Results. *Transportation Research Part B,* Vol. 36, No. 3, pp. 225–252.

Chen, K., L. Yu, J. Guo, and H. Wen. 2007. Characteristics Analysis of Road Network Reliability in Beijing Based on the Data Logs from Taxis. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Chiao, K.-A., and P. Vovsha. 2006. Bringing Activity-Based Modeling Concept to Technical Maturity: NYMTC Model Experience. Presented at the TMIP Conference, Austin, TX.

Chung, K., J. Rudjanakanoknad, and M. J. Cassidy. 2007. Relation Between Traffic Density and Capacity Drop at Three Freeway Bottlenecks. *Transportation Research Part B,* Vol. 41, No. 1, pp. 82–95.

Cirillo, C., and K. W. Axhausen. 2006. Evidence on the Distribution of Values of Travel Time Savings from a Six-Week Travel Diary. *Transportation Research Part A,* Vol. 40, No. 5, pp. 444–457.

Clark, S. D., and D. P. Watling. 2005. Modeling Network Travel Time Reliability Under Stochastic Demand. *Transportation Research Part B,* Vol. 39, No. 2, pp. 119–140.

Cohan, H., and F. Southworth. 1999. On the Measurement and Valuation of Travel Time Variability Due to Incidents on Freeways. *Journal of Transportation and Statistics,* Vol. 2, Dec., pp. 123–131.

Concas, S., and A. Kolpakov. 2009. *Synthesis of Research on Value of Time and Value of Reliability.* Report BD 549–37. Florida Department of Transportation, Tallahassee.

Copley, G., P. Murphy, and D. Pearce. 2002. Understanding and Valuing Journey Time Variability. Presented at European Transport Conference, Cambridge, United Kingdom.

Daganzo, C. F. 1999. *A Behavioral Theory of Multi-Lane Traffic Flow, Part I: Long Homogeneous Freeway Sections.* Institute of Transportation Studies, University of California, Berkeley.

Daganzo, C. F., and Y. Sheffi. 1977. On Stochastic Models of Traffic Assignment. *Transportation Science,* Vol. 11, No. 3, pp. 253–274.

Dehghani, Y., T. Adler, T. Doherty, and R. Fox. 2003. Development of a New Toll Mode Choice Modeling System for Florida's Turnpike Enterprise. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1858,* Transportation Research Board of the National Academies, Washington, D.C., pp. 9–17.

de Jong, G., M. Kouwenhoven, Y. Tseng, E. Verhoef, and J. Bates. 2007. *The Value of Travel Time and Travel Time Reliability: Survey Design.* Final report. Dutch Ministry of Transport, Public Works, and Water Management, Leiden, Netherlands.

de Jong, G. C., E. Kroes, R. Plasmeijer, P. Sanders, and P. Warffemius. 2004. The Value of Reliability. Presented at European Transport Conference, Strasbourg, France.

De Palma, A., and N. Picard. 2005. Route Choice Decision Under Travel Time Uncertainty. *Transportation Research Part A,* Vol. 39, No. 4, pp. 295–324.

Dial, R. B. 1997. Bicriterion Traffic Assignment: Efficient Algorithms Plus Examples. *Transportation Research Part B,* Vol. 31, No. 5, pp. 357–379.

Dong, J., and H. S. Mahmassani. 2009a. Flow Breakdown, Travel Reliability and Real-Time Information in Route Choice Behavior. In *Proceedings of the 18th International Symposium on Transportation and Traffic Theory* (W. Lam, S. C. Wong, and H. Lo, eds.), Springer, pp. 675–695.

Dong, J., and H. S. Mahmassani. 2009b. Flow Breakdown and Travel Time Reliability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2124,* Transportation Research Board of the National Academies, Washington, D.C., pp. 203–212.

Du, Z. P., and A. Nicholson. 1997. Degradable Transportation Systems: Sensitivity and Reliability Analysis. *Transportation Research Part B,* Vol. 31, No. 3, pp. 225–237.

Elefteriadou, L., and X. Cui. 2007. Travel Time Reliability and Truck Level of Service on the Strategic Intermodal System Part A: Travel Time Reliability. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Elefteriadou, L., R. P. Roess, and W. R. McShane. 1995. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record 1484,* TRB, National Research Council, Washington, D.C., pp. 80–89.

Eliasson, J. 2004. Car Drivers' Valuation of Travel Time Variability, Unexpected Delays and Queue Driving. Presented at European Transport Conference, Strasbourg, France.

Emmerink, R. H. M., K. W. Axhausen, P. Nijkamp, and P. Rietveld. 2005. The Potential of Information Provision in a Simulated Road Network with Non-Recurrent Congestion. *Transportation Research C,* Vol. 3, No. 5, pp. 293–309.

Engelson, L. 2011. Properties of Expected Travel Cost Function with Uncertain Travel Time. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2254,* Transportation Research Board of the National Academies, Washington, D.C., pp. 151–159.

Erhardt, G. D., F. S. Koppelman, J. Freedman, W. A. Davidson, and A. Mullins. 2003. Modeling the Choice to Use Toll and HOV Facilities. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1854,* Transportation Research Board of the National Academies, Washington, D.C., pp. 135–143.

Evans, J. L., L. Elefteriadou, and N. Gautam. 2001. Probability of Breakdown at Freeway Merges Using Markov Chains. *Transportation Research Part B,* Vol. 35, No. 3, pp. 237–254.

Florida DOT. 2000. Florida Reliability Method in Florida's Mobility Performance Measures Program. Office of the State Transportation Planner, Transportation Statistics Office, Tallahassee.

Fosgrau, M. 2007. The Value of Reliability and the Distribution of Random Durations. *Proc., European Transport Conference,* Leiden, Netherlands.

Fowkes, A. S., P. E. Firmin, A. E. Whiteing, and G. Tweddle. 2001. Freight Road User Valuations of Three Different Aspects of Delay. *Proc., European Transport Conference,* Cambridge, United Kingdom.

Gibb, J., and J. L. Bowman. 2007. Convergence of an Activity-Based Travel Model System to Equilibrium: Experimental Designs and Findings. Presented at 11th TRB National Planning Applications Conference, Daytona Beach, Fla.

Golob, J. M., and T. F. Golob. 2001. Studying Road Pricing Policy with Panel Data Analysis: The San Diego I-15 HOT Lanes. In *Travel Behavior Research: The Leading Edge,* Pergamon, Oxford, United Kingdom, pp. 869–883.

Golob, T. F. 2001. Joint Models of Attitude and Behavior in the Evaluation of the San Diego I-15 Pricing Project. *Transportation Research Part A,* Vol. 35, No. 6, pp. 495–514.

Golob, T. F., R. Kitamura, and J. Supernak. 1997. A Panel-Based Evaluation of the San Diego I-15 Carpool Lanes Project. In *Panels for Transportation Planning: Methods and Applications* (T. F. Golob, R. Kitamura, and L. Long, eds.), Boston, Mass., Kluwer Academic Publishers, pp. 97–128.

Gordon, A., T. Van Vuren, D. P. Watling, J. Polak, R. B. Noland, S. Porter, and N. Taylor. 2001. Incorporating Variable Traffic Time Effects into Route Choice Models. Proc., *PTRC European Transport Conference* (Methodological Innovations), PTRC Education & Research Services Ltd., London.

Hacura, A., M. Jadamus-Hacura, and A. Kocot. 2001. Risk Analysis in Investment Appraisal Based on the Monte Carlo Simulation Technique. *European Physical Journal B,* Vol. 20, No. 4, pp. 551–553.

Hagen, L., H. Zhou, and F. Pirinccioglu. 2006. Development of Revised Methodology for Collecting Origin-Destination Data. Technical report. Center for Urban Transportation Research, University of South Florida, Tampa.

Hamdar, S. H. 2009. *Modeling Driver Behavior As a Stochastic Hazard-Based Risk-Taking Process.* PhD dissertation. Northwestern University, Evanston, Ill.

Hamdar, S. H., and H. S. Mahmassani. 2008. From Existing Accident-Free Car-Following Models to Colliding Vehicles: Exploration and Assessment. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2088,* Transportation Research Board of the National Academies, Washington, D.C., pp. 45–56.

Hamdar, S. H., M. Treiber, H. S. Mahmassani, and A. Kesting. 2008. Modeling Driver Behavior As a Sequential Risk-Taking Task. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2088,* Transportation Research Board of the National Academies, Washington, D.C., pp. 208–217.

Hensher, D. 2007. Joint Estimation of Process and Outcome in Choice Experiments Involving Attribute Framing. Working paper ITLS-WP-07-04. Institute of Transport and Logistics Studies, University of Sydney, Australia.

Hensher, D. A., and P. Goodwin. 2003 Using Values of Travel Time Savings for Toll Roads: Avoiding Some Common Errors. *Transport Policy,* Vol. 11, No. 2, pp. 171–181.

Hensher, D., and W. H. Greene. 2000. The Mixed Logit Model: The State of Practice. *Transportation,* Vol. 30, No. 2, pp. 133–176.

Hess, S., M. Bierlaire, and J. Polak. 2005. Estimation of Value of Time Using Mixed Logit Models. *Transportation Research Part A,* Vol. 39, No. 3, pp. 221–236.

Hess, S., Polak, J. W., A. Daly, and G. Hyman. 2007. Flexible Substitution Patterns in Models of Mode and Time of Day Choice: New Evidence from the UK and the Netherlands. *Transportation,* Vol. 34, No. 2, pp. 213–238.

Highway Capacity Manual. 2000. TRB, National Research Council, Washington, D.C.

Holguín-Veras, J., K. Ozbay, and A. L. C. de Cerreño. 2005. *Evaluation Study of Port Authority of New York and New Jersey's Time of Day*

Pricing Initiative. Final report. Federal Highway Administration, U.S. Department of Transportation; New Jersey Department of Transportation; and Port Authority of New York and New Jersey.

Hollander, Y. 2006. Direct versus Indirect Models for the Effect of Unreliability. *Transportation Research Part A,* Vol. 40, No. 9, pp. 699–711.

Jackson, W. B., and J. V. Jucker. 1982. An Empirical Study of Travel Time Variability and Travel Choice Behavior. *Transportation Science,* Vol. 16, No. 4, pp. 460–475.

Jayakrishnan, R., H. S. Mahmassani, and T.-Y. Hu. 1994. An Evaluation Tool for Advanced Traffic Information and Management Systems in Urban Networks. *Transportation Research Part C: Emerging Technologies,* Vol. 2, No. 3, Elsevier, pp. 129–147.

Jones, E. G. 1988. *Characterizing Travel Time Variability in a Commuting Corridor.* MS thesis. University of Texas at Austin.

Kaparias, I., M. G. H. Bell, and H. Belzner. 2008. A New Measure of Travel Time Reliability for In-Vehicle Navigation Systems. *Journal of Intelligent Transportation Systems,* Vol. 12, No. 4, pp. 202–211.

Kerner, B. S. 2000. Theory of Breakdown Phenomenon at Highway Bottlenecks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1710,* TRB, National Research Council, Washington, D.C., pp. 136–144.

Kerner, B. S. 2004. *The Physics of Traffic.* Springer, Heidelberg, Germany.

Kim, H., J.-S. Oh, and R. Jayakrishnan. 2006. Activity Chaining Model Incorporating Time Use Problem and Its Application to Network Demand Analysis. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C.

Kitamura, R., and J. Suparnak. 1997. Temporal Utility Profiles of Activities and Travel: Some Empirical Evidence. In *Understanding Travel Behavior in an Era of Change* (P. Stopher and M. Lee-Gosselin, eds.), Elsevier, pp. 339–350.

Kouwenhoven, M., A. Schoemakers, R. van Grol, and E. Kroes. 2005. Development of a Tool to Assess the Reliability of Dutch Road Networks. Presented at European Transport Conference, Strasbourg, France.

Kulkarni, V. G. 1995. *Modeling and Analysis of Stochastic Systems.* Chapman & Hall, London.

Lam, T. 2004. Evaluating Value-Pricing Projects with Both Scheduling and Route Choices. *Regional Science and Urban Economics,* Vol. 34, No. 2, pp. 225–240.

Lam, T. C., and K. A. Small. 2001. The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E,* Vol. 37, No. 2–3, pp. 231–251.

Lam, W., and Y. Yin. 2001. An Activity-Based Time-Dependent Traffic Assignment Model. *Transportation Research Part B,* Vol. 35, No. 6, pp. 549–574.

Leurent, F. 1993. Cost Versus Time Equilibrium Over a Network. *European Journal of Operations Research,* Vol. 71, pp. 205–221.

Levinson, D., K. Harder, J. Bloomfield, and K. Winiarczyk. 2004. Weighting Waiting: Evaluating Perception of In-Vehicle Travel Time Under Moving and Stopped Conditions. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1898,* Transportation Research Board of the National Academies, Washington, D.C., pp. 61–68.

Li, R., G. Rose, and M. Sarvi. 2006. Using Automatic Vehicle Identification Data to Gain Insight into Travel Time Variability and Its Causes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1945,* Transportation Research Board of the National Academies, Washington, D.C., pp. 24–32.

Li, Z., D. A. Hensher, and J. M. Rose. 2010. Willingness to Pay for Travel Time Reliability in Passenger Transport: A Review and Some New Empirical Evidence. *Transportation Research Part E,* Vol. 46, No. 3, pp. 384–403.

Liu, X., and J. W. Polak. 2007. Nonlinearity and Specification of Attitudes Toward Risk in Discrete Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, Washington, D.C., pp. 27–31.

Lomax, T., D. Schrank, S. Turner, and R. Margiotta. 2003. *Selecting Travel Reliability Measures.* Federal Highway Administration, U.S. Department of Transportation.

Lu, C.-C., and H. S. Mahmassani. 2007. Heterogeneous Users and Variable Road Pricing: Model and Algorithm for the Bi-Criterion Dynamic User Equilibrium Problem on Large Networks. Presented at 6th Triennial Symposium on Transportation Analysis (TRISTAN VI), Phuket Island, Thailand.

Lu, C.-C., H. S. Mahmassani, and X. Zhou. 2006. Variable Toll Pricing and Heterogeneous Users: Model and Solution Algorithm for Bicriterion Dynamic Traffic Assignment Problem. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1964,* Transportation Research Board of the National Academies, Washington, D.C., pp. 19–26.

Lu, C.-C., H. S. Mahmassani, X. Zhou. 2007. A Bi-criterion Dynamic User Equilibrium Traffic Assignment Model and Solution Algorithm for Evaluating Dynamic Road Pricing Strategies. *Transportation Research Part C: Emerging Technologies,* Vol. 16, No. 4, Elsevier, pp. 371–389.

Lu, C.-C., H. S. Mahmassani, X. Zhou. 2009. Equivalent Gap Function-based Reformulation and Solution Algorithm for the Dynamic User Equilibrium Problem. *Transportation Research Part B: Methological,* Vol. 43, No. 3, Elsevier, pp. 345–364.

Lyman, K., and R. L. Bertini. 2008. Using Travel Time Reliability Measures to Improve Regional Transportation Planning and Operations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2046,* Transportation Research Board of the National Academies, Washington, D.C., pp. 1–10.

Mahmassani, H. S. 2001. Dynamic Network Traffic Assignment and Simulation Methodology for Advanced Systems Management Applications. *Networks and Spatial Economics,* Vol. 1, pp. 267–292.

Mahmassani, H. S., and R. Herman. 1987. Interaction of Trip Decisions and Traffic Systems Dynamics. *European Journal of Operational Research,* Vol. 30, No. 3, pp. 304–317.

Mahmassani, H. S., G.-L. Chang, and R. Herman. 1986. Individual Decisions and Collective Effects in a Simulated Traffic System. *Transportation Science,* Vol. 20, No. 4, Elsevier, pp. 258–271.

Mahmassani, H. S., E. Jones, R. Herman, and C. M. Walton. 1989. Travel Time Variability in a Commuting Corridor: Implications for Electronic Route Guidance. *Proc., 1st International Conference on Application of Advanced Technologies in Transportation Engineering,* American Society of Civil Engineers, San Diego, Calif.

Mahmassani, H. S., and C. Tong. 1986. Availability of Information and Dynamics of Departure Time Choice: Experimental Investigation. *Transportation Research Record: Journal of the Transportation Research Board No. 1085,* Transportation Research Board of the National Academies, Washington, D.C., pp. 33–49.

Mahmassani, H. S., X. Zhou, and C. Lu. 2005. Toll Pricing and Heterogeneous Users: Approximation Algorithms for Finding Bi-Criterion Time-Dependent Efficient Paths in Large-Scale Traffic Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1923,* Transportation Research Board of the National Academies, Washington, D.C., pp. 28–36.

Marcotte, P. 1999. Reformulations of a Bi-Criterion Equilibrium Model. In *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and*

Smoothing Methods (M. Fukushima and L. Qi, eds.), Kluwer Academic, Dordrecht, Netherlands, pp. 269–292.

Marcotte, P., and D. L. Zhu. 1997. Equilibria with Infinitely Many Differential Classes of Customers. In *Complementarity and Variational Problems: State of the Art. Proceedings of the 13th International Conference on Complementarity Problems* (J.-S. Pang and M. Ferris, eds.), SIAM, Philadelphia, Pa., pp. 234–258.

Martin, W. A., and N. A. McGuckin. 1998. *NCHRP Report 365: Travel Estimation Techniques for Urban Planning.* TRB, National Research Council, Washington, D.C.

Mastako, K. A. 2003. *Choice Set As an Indicator for Choice Behavior When Lanes Are Managed with Value Pricing.* PhD dissertation. Texas A & M University, College Station.

McCormick Rankin Corp., and Parsons Brinckerhoff. 2008. *TRANS Model Redevelopment: City of Ottawa.* Technical report. Ottawa, Ontario, Canada.

McNeill, D. R. 1968. A Solution to the Fixed-Cycle Traffic Light Problem for Compound Poisson Arrivals. *Journal of Applied Probability,* Vol. 5, No. 3, pp. 624–635.

Michea, A., and J. W. Polak. 2006. Modeling Risky Choice Behavior: Evaluating Alternatives to Expected Utility Theory. Presented at 11th International Conference on Travel Behavior Research, Kyoto, Japan.

Mirchandani, P., and H. Soroush. 1987. Generalized Traffic Equilibrium with Probabilistic Travel Times and Perceptions. *Transportation Science,* Vol. 21, No. 3, pp. 133–152.

National Cooperative Highway Research Program (NCHRP). 2009. *Improved Framework and Tools for Highway Pricing Decisions.* NCHRP 08-57. Federal Highway Administration (FHWA), Washington, D.C.

Noland, R. B., and J. W. Polak. 2002. Travel Time Variability: A Review of Theoretical and Empirical Issues. *Transport Reviews,* Vol. 22, No. 1, pp. 39–54.

Noland, R. B., and K. A. Small. 1995. Travel Time Uncertainty, Departure Time Choice and the Cost of the Morning Commute. Working paper UCI-ITS-WP-95-1. Institute of Transportation Studies, University of California, Irvine.

Noland, R. B., K. A. Small, P. M. Koskenoja, and X. Chu. 1998. Simulating Travel Reliability. *Regional Science & Urban Economics,* Vol. 28, No. 5, pp. 535–564.

Palmquist, R., D. Phaneuf, and K. Smith. 2007. Measuring the Values for Time. NBER Working Paper No. 13594, National Bureau of Economic Research, Cambridge, Mass.

Parsons Brinckerhoff Quade and Douglas, Inc., PB Consult, Inc., AECOM Consult, Urbitran Associates, Inc., Urbanomics, Alex Anas and Associates, NuStats International, and George Hoyt and Associates, Inc. 2005. *Transportation Models and Data Initiative: General Report: New York Best Practice Model (NYBPM).* New York Metropolitan Transportation Council. U.S. Department of Transportation.

Payne, H. J. 1984. Discontinuity in Equilibrium Freeway Traffic Flow. In *Transportation Research Record 971,* TRB, National Research Council, Washington, D.C., pp. 140–146.

PB Consult, Inc. 2003. *Travel Demand Model Development for Traffic and Revenue Studies for Public-Private Partnership Highway Projects in the Montreal Region.* Ministry of Transportation of Québec, Québec, Canada.

PB Consult, Inc. 2005. *MORPC Travel Demand Model: Validation and Final Report.* Mid-Ohio Regional Planning Commission, Columbus.

Persaud, B., S. Yagar, and R. Brownlee. 1998. Exploration of the Breakdown Phenomenon in Freeway Traffic. *Transportation Research Record 1634,* TRB, National Research Council, Washington, D.C., pp. 64–69.

Polak, J. W., S. Hess, and X. Liu. 2008. Characterizing Heterogeneity in Attitudes to Risk in Expected Utility Models of Mode and Departure Time Choice. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C.

Popuri, Y., M. E. Ben-Akiva, K. Proussalogou. 2008. Time-of-Day Modeling in a Tour-Based Context: Tel Aviv Experience. *Transportation Research Record: Journal of the Transportation Research Board, No. 2076,* Transportation Research Board of the National Academies, Washington, D.C., pp. 88–96.

Puget Sound Regional Council. 2008. Traffic Choices Study: Summary Report. FHWA, U.S. Department of Transportation. www.psrc.org/assets/37/summaryreport.pdf. Accessed March 11, 2013.

Rakha, H., I. El-Shawarby, M. Arafeh, and F. Dion. 2006. Estimating Path Travel Time Reliability. In *ITSC 2006: 9th International IEEE Conference on Intelligent Transportation Systems,* Toronto, Ontario, Canada, pp. 236–241.

Rakha, H., and Z. Wang. 2005. Estimating Traffic Stream Space Mean Speed and Reliability from Dual- and Single-Loop Detectors. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1925,* Transportation Research Board of the National Academies, Washington, D.C., pp. 28–47.

Recker, W., Y. S. Chung, J. Y. Park, L. Wang, A. Chen, Z. Ji, H. Liu, M. Horrocks, and J. S. Oh. 2005. *Considering Risk-Taking Behavior in Travel Time Reliability.* California Partners for Advanced Transit and Highway (PATH) research report. Institute of Transportation Studies, University of California, Berkley.

Redmond, L., and P. Mokhtarian. 2001. The Positive Utility of the Commute: Modeling Ideal Commute Time and Relative Desired Commute Amount, Transportation, Vol. 28, No. 2, pp. 179–205.

Ritsema van Eck, J., D. Snellen, and H. Hilbers. 2004. Travel Time Reliability: Methodology and Some Results for the Netherlands. Presented at World Conference on Transportation Research, Istanbul, Turkey.

Sbayti, H., C. C. Lu, H. S. Mahmassani. 2007. Efficient Implementation of Method of Successive Averages in Simulation-Based Dynamic Traffic Assignments Models for Large-Scale Network Applications. *Transportation Research Record: Journal of the Transportation Research Board, No. 2029,* Transportation Research Board of the National Academies, Washington, D.C., pp. 22–30.

Schönhof, M., and D. Helbing, D. 2007. Empirical Features of Congested Traffic States and Their Implications for Traffic Modeling. *Transportation Science,* Vol. 41, No. 2, pp. 135–166.

Senna, L. A. D. S. 1991. Risk of Delays, Uncertainty, and Travelers' Valuation of Travel Time Variability. *Proc., 19th PTRC Summer Annual Meeting,* Brighton, United Kingdom.

Sillano, M., and J. Dios de Ortúzar. 2005. Willingness-to-Pay Estimation with Mixed Logit Models: Some New Evidence. *Environment and Planning A,* Vol. 37, No. 3, pp. 525–550.

Slavin, H., J. Brandon, and A. Rabinowicz. 2007. Empirical Comparison of Alternative Equilibrium Assignment Methods. Presented at 11th TRB National Planning Applications Conference, Daytona Beach, Fla.

Small, K. A. 1982. The Scheduling of Consumer Activities: Work Trips. *American Economic Review,* Vol. 72, No. 3, pp. 467–479.

Small, K. A. 1999. Project Evaluation. In *Essays in Transportation Economics and Policy: A Handbook in Honor of John Meyer* (J. A. Gomez-Ibanez, W. B. Tye, and C. Winston, eds.), Brookings Institution, Washington, D.C., pp. 137–177.

Small, K. A., R. Noland, X. Chu, and D. Lewis. 1999. *NCHRP Report 431: Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation.* TRB, National Research Council, Washington, D.C.

Small, K. A., and H. S. Rosen. 1981. Applied Welfare Economics with Discrete Choice Models. *Econometrica,* Vol. 49, No. 1, pp. 105–130.

Small, K. A., and E. T. Verhoef. 2007. *Economics of Urban Transportation.* Routledge, London and New York.

Small, K. A., C. Winston, and J. Yan. 2005. Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica,* Vol. 73, No. 4, pp. 1367–1382.

Small, K. A., C. Winston, and J. Yan. 2006. Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design. Brookings-Wharton Papers on Urban Affairs, Brookings Institution, Washington, D.C.

Sohn, K., and D. Kim. 2009. Statistical Model for Forecasting Link Travel Time Variability. *Journal of Transportation Engineering,* Vol. 135, No. 7, pp. 440–453.

Songchitruksa, P., and K. Balke. 2006. Assessing Weather, Environment, and Loop Data for Real-Time Freeway Incident Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1959,* Transportation Research Board of the National Academies, Washington, D.C., pp. 105–113.

Sorensen, P. A., and B. D. Taylor. 2006. Innovations in Road Finance: Examining the Growth in Electronic Tolling. *Public Works Management & Policy,* Vol. 11, No. 2, pp. 110–125.

Spear, B. D. 2005. A Summary of the Current State of the Practice in Modeling Road Pricing. Presented at Expert Forum on Road Pricing and Travel Demand Modeling, U.S. Department of Transportation, Alexandria, Va.

Stefan, K., J. McMillan, C. Blaschuk, and J. Hunt J. 2007. Estimation of a Weekend Location Choice Model for Calgary. Presented at 11th Transportation Planning Application Conference, TRB, Daytona Beach, Fla.

Stephan, D. G., and H. S. Mahmassani. 1988. Experimental Investigation of Route and Departure Time Choice Dynamics of Urban Commuters. *Transportation Research Record: Journal of the Transportation Research Record No. 1203.* Transportation Research Board of the National Academies, Washington, D.C., pp. 64–84.

Steimetz, S. C., and D. Brownstone. 2005. Estimating Commuters' "Value of Time" with Noisy Data: A Multiple Imputation Approach. *Transportation Research Part B,* Vol. 39, No. 10, pp. 865–889.

Stockton, W., R. Benz, L. Rilett, D. Skowronek, S. Vadali, and G. Daniels. 2000. *Investigation of General Feasibility of High Occupancy/Toll Lanes in Texas.* TTI Report 4915-1. Texas Transportation Institute, Texas A&M University, College Station.

Stockton, W., P. Hughes, M. Hickman, D. Pucket, Q. Brown, A. Miranda, and S. Woong. 2000. *An Evaluation of the Katy Freeway HOV Lane Pricing Project.* TTI Report E 205001. Texas Transportation Institute, Texas A&M University, College Station.

Sullivan, E. C. 2000. *Continuation Study to Evaluate the Impacts of the SR 91 Value-Priced Express Lanes.* Final report. California Department of Transportation, Sacramento.

Sullivan, E. C. 2002. State Route 91 Value-Priced Express Lanes: Updated Observations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1812,* Transportation Research Board of the National Academies, Washington, D.C., pp. 37–42.

Sumalee, A., and D. Watling. 2008. Partition-Based Algorithm for Estimating Transportation Network Reliability with Dependent Link Failures. *Journal of Advanced Transportation,* Vol. 42, No. 3, pp. 213–238.

Supernak, J. 1992. Temporal Utility Profiles of Activities and Travel: Uncertainty and Decision Making. *Transportation Research Part B,* Vol. 26, Vol. 1, pp. 549–574.

Tilahun, N. Y., and D. M. Levinson. 2007. Value of Time Comparisons in the Presence of Unexpected Delay. Presented at 3rd International Conference on Transport Network Reliability, Delft, Netherlands.

Tilahun, N. Y., and D. M. Levinson. 2010. A Moment of Time: Reliability in Route Choice Using Stated Preference. *Journal of Intelligent Transportation Systems,* Vol. 14, No. 3, pp. 179–187.

Tillman, R., J. Smolley, K. Massarelli, A. Goldberg, A. Pratt, and P. Eshelman. 2006. Long-Term Projection of Traffic and Revenues for Equity Analysis. *Tollways: Focus on Finance,* Autumn, International Bridge, Tunnel, and Turnpike Association.

Train, K., and G. Sonnier. 2004. Mixed Logit with Bounded Distributions of Correlated Partworths. In *Applications of Simulation Methods in Environmental and Resource Economics* (R. Scarpa and A. Alberini, eds.), Kluwer Academic Publishers, Boston, Mass.

Tseng, Y. Y., B. Ubbels, and E. Verhoef. 2005. Value of Time, Schedule Delay and Reliability: Estimation Results of a Stated Choice Experiment Among Dutch Commuters Facing Congestion. European Regional Science Association (ERSA) Conference.

Tseng, Y., and E. T. Verhoef. 2008. Value of Time by Time of Day: A Stated Preference Study. *Transportation Research Part B,* Vol. 42, No. 7–8, pp. 607–618.

Tseng, Y., E. T. Verhoef, and A. I. J. M. van der Hoorn. 2007. A Pilot Study into the Perception of Unreliability of Travel Time Using In-Depth Interviews. Presented at European Transport Conference, Noordwijk, Netherlands.

Tu, H., H. van Lint, and H. van Zuylen. 2005. Real-Time Modeling Travel Time Reliability on Freeway. Presented at 10th Euro Working Group Transportation Meeting and 16-Mini-Euro Conference, Poznan, Poland.

Uchida, T., and Y. Iida. 1993. Risk Assignment: A New Traffic Assignment Model Considering Risk of Travel Time Variation. In *Proc., 12th International Symposium on Transportation and Traffic Theory,* Elsevier, Amsterdam, pp. 89–105.

van den Broek, M. S., J. S. H. van Leeuwaarden, I. J. B. F. Adan, and O. J. Boxma. 2004. Bounds and Approximations for the Fixed-Cycle Traffic-Light Queue. *SPOR Report,* Series 18.

van Lint, J. W., and H. J. van Zuylen. 2005. Monitoring and Predicting Freeway Travel Time Reliability Using Width and Skew of Day-to-Day Travel Time Distribution. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1917,* Transportation Research Board of the National Academies, Washington, D.C., pp. 54–62.

van Lint, J. W., H. J. van Zuylen, and H. Tu. 2008. Travel Time Unreliability on Freeways: Why Measures Based on Variance Tell Only Half of the Story. *Transportation Research Part A,* Vol. 42, No. 1, pp. 258–277.

Verhoef, E. T. 2007. Second-Best Road Pricing Through Highway Franchising. *Journal of Urban Economics,* Vol. 62, No. 2, pp. 337–361.

Verhoef, E. T., and K. A. Small. 2004. Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users. *Journal of Transport Economics and Policy,* Vol. 38, No. 1, pp. 127–156.

Vollmer Associates. 2001. *Northwest Parkway Traffic and Revenue Study.* Northwest Parkway Public Highway Authority, Denver, Colo.

Vovsha, P., and M. Bradley. 2004. Hybrid Discrete Choice Departure-Time and Duration Model for Scheduling Travel Tours. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894,* Transportation Research Board of the National Academies, Washington, D.C., pp. 46–56.

Vovsha, P., M. Bradley, and J. Bowman. 2005. Activity-Based Travel Forecasting Models in the United States: Progress Since 1995 and Prospects for the Future. In *Progress in Activity-Based Analysis* (H. Timmermans, ed.), Elsevier, pp. 389–414.

Vovsha, P., W. Davidson, and R. Donnelly. 2005. Making the State of the Art the State of the Practice: Advanced Modeling Techniques for Road Pricing. Presented at Expert Forum on Road Pricing and

Travel Demand Modeling, U.S. Department of Transportation, Alexandria, Va.

Vovsha, P., R. Donnelly, and K. Chiao. 2008. Activity-Based Model for the Mega-Region of New York. *Traffic Engineering and Control.* Vol. 49, No. 9, pp. 336–340.

Vovsha, P., R. Donnelly, and S. Gupta. 2008. Network Equilibrium with Activity-Based Microsimulation Models: The New York Experience. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2054,* Washington, D.C., pp. 102–109.

Vovsha, P., S. Gupta, J. Freedman, W. Sun, and V. Livshits. 2012. Workplace Choice Model: Comparison of Spatial Patterns of Commuting in Four Metropolitan Regions. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C.

Vovsha, P., and E. Petersen. 2005. Escorting Children to School: Statistical Analysis and Applied Modeling Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1921,* Transportation Research Board of the National Academies, Washington, D.C., pp. 131–140.

Vovsha, P., E. Petersen, and R. M. Donnelly. 2003. Explicit Modeling of Joint Travel by Household Members: Statistical Evidence and Applied Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1831,* Transportation Research Board of the National Academies, Washington, D.C., pp. 1–10.

Vovsha, P., E. Petersen, R. M. Donnelly. 2002. Micro-Simulation in Travel Demand Modeling: Lessons Learned from the New York Best Practice Model. *Transportation Research Record: Journal of the Transportation Research Board, No. 1805,* Transportation Research Board of the National Academies, Washington, D.C., pp. 68–77.

Wardman, M. 1998. The Value of Travel Time: A Review of British Evidence. *Journal of Transport Economics and Policy,* Vol. 32, No. 3, pp. 285–316.

Wardman, M., J. N. Ibanez, and J. Pagan. 2009. Variations in Motorists' Valuations of Travel Time with Traffic Conditions. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Warffemius, P. 2005. Using the Standard Deviation of the Travel Time Distribution as an Indicator for Valuing the Reliability of Travel Time. Presented at European Conference of Transport Research Institutes, The Hague, Netherlands.

Watling, D., A. Sumalee, R. Connors, and C. Balijepalli. 2004. *Advancing Methods for Evaluating Network Reliability.* Final report. Department for Transport New Horizons Study, United Kingdom.

Watling, D. P. 2006. User Equilibrium Traffic Network Assignment with Stochastic Travel Times and Late Arrival Penalty. *European Journal of Operational Research,* Vol. 175, No. 3, pp. 1539–1556.

Yan, J., and K. A. Small. 2002. Choice Models of Route, Occupancy, and Time of Day with Value-Priced Tolls. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1812,* Transportation Research Board of the National Academies, Washington, D.C., pp. 69–77.

Yang, H., W. H. Tang, W. M. Cheung, and Q. Meng. 2002. Profitability and Welfare Gain of Private Toll Roads in a Network with Heterogeneous Users. *Transportation Research Part A,* Vol. 36, No. 6, 2002, pp. 537–554.

Yin, Y., and H. Ieda. 2001. Assessing Performance Reliability of Road Networks Under Recurrent Congestion. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C.

Yin, Y., W. H. K. Lam, and H. Ieda. 2002. Modeling Risk-Taking Behavior in Queuing Networks with Advanced Traveler Information Systems. In *Transportation and Traffic Theory in the 21st Century* (M.A.P. Tailor, ed.), Elsevier, Oxford, pp. 309–328.

Zhang, K., H. S. Mahmassani, and C.-C. Lu. 2008. Probit-based Time-dependent Stochastic User Equilibrium Traffic Assignment Model. *Transportation Research Record: Journal of the Transportation Research Board, No. 2085,* Transportation Research Board of the National Academies, Washington, D.C., pp. 86–94.

Zhang, N., and W. H. K. Lam. 2001. An Alternative to the Road Network Reliability by Travel Demand Satisfaction Ratio. In *Proc., 4th International Conference on Management,* Xian Jiaotong University, China, pp. 47–53.

Zhou, X., H. S. Mahmassani, and K. Zhang. 2008. Dynamic Micro-Assignment Modeling Approach for Integrated Multimodal Urban Corridor Management. *Transportation Research Part C,* Vol. 16, No. 2, pp. 167–186.

TRB OVERSIGHT COMMITTEE FOR THE STRATEGIC HIGHWAY RESEARCH PROGRAM 2*

CHAIR: **Kirk T. Steudle,** *Director, Michigan Department of Transportation*

MEMBERS

H. Norman Abramson, *Executive Vice President (retired), Southwest Research Institute*
Alan C. Clark, *MPO Director, Houston–Galveston Area Council*
Frank L. Danchetz, *Vice President, ARCADIS-US, Inc.*
Malcolm Dougherty, *Director, California Department of Transportation*
Stanley Gee, *Executive Deputy Commissioner, New York State Department of Transportation*
Mary L. Klein, *President and CEO, NatureServe*
Michael P. Lewis, *Director, Rhode Island Department of Transportation*
John R. Njord, *Executive Director (retired), Utah Department of Transportation*
Charles F. Potts, *Chief Executive Officer, Heritage Construction and Materials*
Ananth K. Prasad, *Secretary, Florida Department of Transportation*
Gerald M. Ross, *Chief Engineer (retired), Georgia Department of Transportation*
George E. Schoener, *Executive Director, I-95 Corridor Coalition*
Kumares C. Sinha, *Olson Distinguished Professor of Civil Engineering, Purdue University*
Paul Trombino III, *Director, Iowa Department of Transportation*

EX OFFICIO MEMBERS

Victor M. Mendez, *Administrator, Federal Highway Administration*
David L. Strickland, *Administrator, National Highway Transportation Safety Administration*
Frederick "Bud" Wright, *Executive Director, American Association of State Highway and Transportation Officials*

LIAISONS

Ken Jacoby, *Communications and Outreach Team Director, Office of Corporate Research, Technology, and Innovation Management, Federal Highway Administration*
Tony Kane, *Director, Engineering and Technical Services, American Association of State Highway and Transportation Officials*
Jeffrey F. Paniati, *Executive Director, Federal Highway Administration*
John Pearson, *Program Director, Council of Deputy Ministers Responsible for Transportation and Highway Safety, Canada*
Michael F. Trentacoste, *Associate Administrator, Research, Development, and Technology, Federal Highway Administration*

*Membership as of April 2013.

CAPACITY TECHNICAL COORDINATING COMMITTEE*

CHAIR: **Mark Van Port Fleet,** *Director, Bureau of Highway Development, Michigan Department of Transportation*

MEMBERS

Kome Ajise, *Program Manager, Public-Private Partnership Program, California Department of Transportation (Caltrans)*
Michael Bruff, *Manager, Transportation Planning Branch, North Carolina Department of Transportation*
Jacquelyn D. Grimshaw, *Vice President for Policy, Center for Neighborhood Technology*
Kris Hoellen, *Director, Conservation Leadership Network, The Conservation Fund*
Carolyn H. Ismart, *Florida Department of Transportation (retired)*
Randy Iwasaki, *Executive Director, Contra Costa Transportation Authority*
Thomas J. Kane, *Thomas J. Kane Consulting*
Keith L. Killough, *Assistant Director, Travel Demand Modeling and Analysis, Multimodal Planning Division, Arizona Department of Transportation*
T. Keith Lawton, *Principal, Keith Lawton Consulting, Inc.*
Debra Nelson, *Strategic Policy Advisor, New York State Department of Transportation*
Bob Romig, *State Transportation Development Administrator, Florida Department of Transportation*
Joseph L. Schofer, *Professor of Civil Engineering and Environmental Engineering and Associate Dean, McCormick School of Engineering and Applied Science, Northwestern University*
Barry Seymour, *Executive Director, Delaware Valley Regional Planning Commission*
Brian J. Smith, *Washington State Department of Transportation*
John V. Thomas, *Office of Policy, Economics, and Innovation, Environmental Protection Agency*
Gary Toth, *Director, Project for Public Spaces*
Jeff Welch, *Director, Knoxville Regional Transportation Planning Organization*
Doug Woodall, *State Director, Turnpike Planning and Development, Texas Turnpike Authority Division, Texas Department of Transportation*

AASHTO LIAISONS

Janet P. Oakley, *Director, Policy and Government Relations, American Association of State Highway and Transportation Officials*
Matthew Hardy, *Program Director, Planning and Policy, American Association of State Highway and Transportation Officials*

FHWA LIAISONS

James Cheatham, *Director, Office of Planning, Office of Planning HEPP-1, Federal Highway Administration*
Gary A. Jensen, *Team Leader, Byways, TCSP & Delta Programs, Office of Human Environment HEPH-30, Federal Highway Administration*
Spencer Stevens, *Community Planner, Office of Planning Oversight and Stewardship, Federal Highway Administration*

*Membership as of February 2013.

Related SHRP 2 Research

Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network (C10A)

Partnership to Develop an Integrated Advanced Travel Demand Model with Mode Choice Capability and Fine-Grained, Time-Sensitive Networks (C10B)

Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools (L04)

Local Methods for Modeling, Economic Evaluation, Justification and Use of the Value of Travel Time Reliability in Transportation Decision Making (L35A and L35B)