The Travel Model Improvement Program

Travel Model Validation and Reasonableness Checking Manual
Second Edition

Helping Agencies Improve Their Planning Analysis Techniques

TMIP

Travel Model Improvement Program

Travel Model Validation and Reasonableness Checking Manual Second Edition

Original: May 14, 2010
Final: September 24, 2010

Prepared by:
Cambridge Systematics, Inc.
100 CambridgePark Drive, Suite 400
Cambridge, Massachusetts 02140

Prepared for:

Federal Highway Administration

Table of Contents

Table of Contents
(continued)

List of Tables

List of Tables
(continued)

List of Tables
(continued)

List of Figures

List of Figures
(continued)

1.0 Introduction

1.0 Introduction

Essentially, all models are wrong, but some are useful…the practical question is how wrong do they have to be to not be useful.[1]

1.1 Why Is Validation Important?

As stated in the introductory quote, "all models are wrong…" An obvious question that might be asked in response to that statement is, "Why?" The answer is relatively simple. Travel models are a closed system, distinguished by a set of mathematical formulae and relationships, being used in an attempt to reflect an open system – the real world populated by people who are responding to influences that are constantly changing, do not always make rational decisions, and whose responses to influences affecting travel are not always the same. Since travel models (and travel modelers) cannot be omniscient, there will always be missed information and abstractions resulting in less than perfect models. Thus, validation is important since it is the effort focused on answering the second part of the quote, the usefulness of the models.

The original *Travel Model Validation and Reasonableness Checking Manual*[2] was published by the Federal Highway Administration (FHWA) Travel Model Improvement Program (TMIP) in 1997. Some of the reasons that manual was produced included:

> "…the lack of attention and effort placed on the validation phase of model development…models need to be able to replicate observed conditions within reason before being used to produce future-year forecasts…credibility of the process with decision-makers will depend largely on the ability of analysts to properly validate procedures and models used.
>
> "…travel modeling process has undergone many changes in the past few years in order to evaluate more complex policy actions…tradeoff between increasing confidence in the level of accuracy of the models and the cost of data collection and effort required to validate models."

---

[1] George Box, Professor Emeritus of Statistics, University of Wisconsin, as quoted in *Project Traffic Forecasting, NCHRP 255 Review*, by Doug Laird, Travel Model Improvement Program (TMIP) Webinar on Project Planning Forecasts, September 25, 2008.

[2] *Travel Model Validation and Reasonableness Checking Manual*, *Travel Model Improvement Program*, Federal Highway Administration, February 1997.

Since the publication of the original manual, the need for quality model validation has increased. Resources for the construction of new facilities have become increasingly scarce, new funding mechanisms such as tolling have become popular, and new travel forecasting techniques focused more on the explanation of traveler behavior than travel patterns have been implemented in a number of regions. These issues have made well validated travel models increasingly important in order to provide decision-makers the assurance they need to confidently use the travel forecasts. However, as documented in the Transportation Research Board (TRB) Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction* (*SR 288*),[3] the state of model validation has not necessarily kept pace with the need (*emphasis added*):

> **Validation Errors:** Validating the ability of a model to predict future behavior requires comparing its predictions with information other than that used in estimating the model. Perceived problems with model validation include *insufficient emphasis and effort* focused on the validation phase, the *unavailability of accurate and current data* for validation purposes, and the *lack of necessary documentation*. The survey of [Metropolitan Planning Organizations] MPOs conducted for this study found that validation is hampered by a *dearth of independent data sources*.

1.2 Background

This manual builds upon several efforts associated with improving the state of model validation and the resulting forecasts:

- The 1990 FHWA publication, *Calibration and Adjustment of System Planning Models*, (FHWA-ED-90-015);
- The 1997 FHWA *Travel Model Validation and Reasonableness Checking Manual*;
- Numerous state publications on travel model validation, such as those by Michigan, Ohio, and Florida;
- The March 11, 2008 TMIP webinar, Shining a Light Inside the Black Box: Model Testing;
- The Federal Transit Administration (FTA) New Starts Workshops and New Starts guidance; and
- The May 9, 2008 TMIP Peer Exchange on Travel Model Validation Practices.

[3] Special Report 288, *Metropolitan Travel Forecasting, Current Practice and Future Direction*, Transportation Research Board (TRB), 2007.

The sources referenced above provide recommendations and guidance regarding travel model validation. However, some of the references are dated, and some have a relatively narrow focus. In addition, new travel forecasting techniques, such as synthesizing regional populations, activity- and tour-based modeling, dynamic traffic assignment, and simulation, are emerging as accepted modeling practices.

The TMIP has provided technical assistance to aid planning organizations in implementing best state-of-the-practice and state-of-the-art modeling methods. As part of the TMIP's efforts to improve travel modeling practice, this manual provides guidance on:

- The development of model validation plans, including collection of proper validation data;
- The role and specification of validation and reasonableness checks and criteria;
- The role of model sensitivity testing in model validation; and
- The development of validation documentation.

This manual focuses on both existing trip-based and emerging activity- and tour-based modeling processes. While it is impossible to specify the checks needed to validate every possible model, this manual describes families of checks and provides concrete examples of their application. The manual is being published as a web-based document with the goal of making it a living document with best practice updates as new issues or techniques emerge.

1.3 Target Audience

This manual has been developed for technical staff charged with the development, maintenance, or application of travel models. Planners, practitioners, policy-level officials, and other stakeholders involved in the transportation process may find the concepts discussed in this manual useful in their evaluation of travel forecasts and the models used to produce those forecasts. The potential target audience includes:

- Travel demand forecasting model development and application staff;
- Planning and operations staff;
- Staff of elected officials and other policy-makers; and
- Staff of nongovernmental organizations concerned with transportation planning or policy issues.

Travel model validation and documentation efforts need to cover a wide range of interests and concerns. A primary goal of this manual is to provide information on validation processes useful for building confidence in the travel modeling process and resulting travel forecasts for the various target audiences. Indeed, at the May 9, 2008 TMIP Peer Exchange, the following definition of was developed for improved model validation *at a level of detail needed to support public decision-making*:

> Steps to verify the ability of the model system to make reasonable predictions over a range of development patterns, transportation operations, and external factors.

1.4 Overview of the Model Validation and Reasonableness Testing Process

1.4.1 Definitions

One confounding issue regarding model validation and reasonableness checking is the lack of a common definition of terms. The task referred to as model validation by one person might be referred to as model calibration by another. Throughout this manual, the following definitions will be employed:

- **Estimation** is the use of statistical analysis techniques and observed data to develop model parameters or coefficients. While model estimation typically occurs at a disaggregate level without bias or correction factors, model estimation may also use statistical analysis procedures to analyze more aggregate data.

- **Assertion** is the declaration of model forms or parameters without the use of statistical analysis of observed data. Model transfer from one region to another is a form of model assertion.

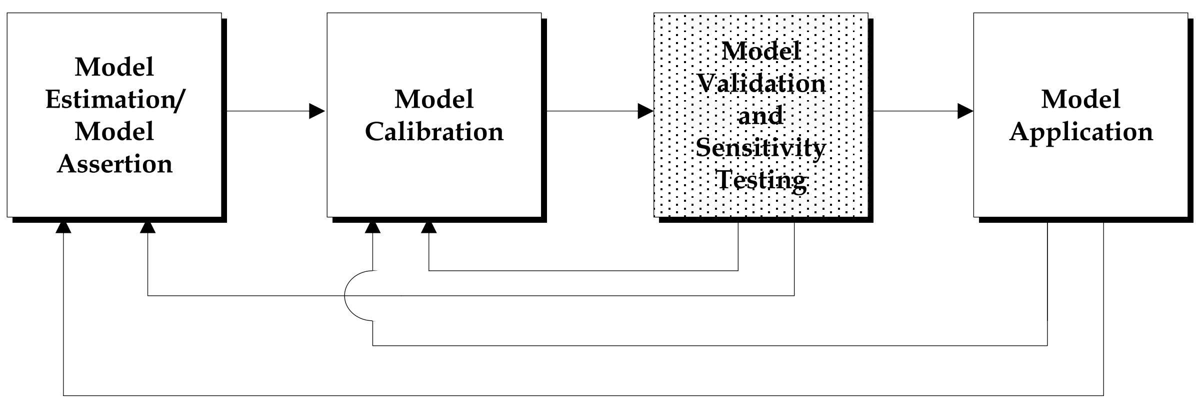
- **Calibration** is the adjustment of constants and other model parameters in estimated or asserted models in an effort to make the models replicate observed data for a base (calibration) year or otherwise produce more reasonable results. Model calibration is often incorrectly considered to be model validation.

- **Validation** is the application of the calibrated models and comparison of the results against observed data. Ideally, the observed data are data *not* used for the model estimation or calibration but, practically, this is not always feasible. Validation data may include additional data collected for the same year as the estimation or calibration of the model or data collected for an alternative year. Validation should also include sensitivity testing.

- **Sensitivity testing** is the application of the models and the model set using alternative input data or assumptions. Sensitivity testing of individual model components may include the estimation of the elasticities and cross-elasticities of model coefficients. However, sensitivity testing should also include the application of the entire model set using alternative assumptions regarding the input demographic data, socioeconomic data, or transportation system to determine if the model results are plausible and reasonable.

The travel model development, validation, and application processes defined above can be viewed as shown in Figure 1.1. Model validation and sensitivity testing may reveal the need to return to the model estimation or model calibration steps. The application of the model using future year conditions and policy options requires checking the reasonableness of projections, and also might reveal a need to return to the model estimation or calibration steps. Issues uncovered during model application never lead directly back to the validation step since it is not possible to improve the model or model forecasts through additional validation. In some cases, however, additional model validation might be helpful in confirming the veracity of forecasts.

Figure 1.1 Overview of Model Development and Application Process



The focus of this manual is the shaded model validation and sensitivity testing task. Some troubleshooting strategies will be provided for situations when satisfactory validations are not obtained. The troubleshooting strategies are applied as part of the model estimation or model calibration steps.

1.4.2 Recommended Model Validation Approach

Validation Process Elements

There are five primary model validation process elements that will be covered in this manual. Each of the elements is the subject of one or more chapters as outlined below:

Validation Process Element	Discussed in Manual Chapter(s)
1. Model Validation Plan Specification	2
2. Collection and Assessment of Validation Data	2
3. Validation of Model Components	3 to 9
4. Validation of Model System	9 to 10
5. Documentation of Validation Results	11

One or more of the above elements is too often skipped, which can lead to inadequate model validation. For example, without a good model validation plan, necessary data may not be available for proper validation of model components or the model system. This can lead to over-reliance on matching observed traffic counts or transit boardings for model validation. Likewise, documentation is often overlooked. Yet quality documentation is key to providing planners, engineers, and decision-makers with a better understanding of the reliability of forecasts and the methods used to reproduce "observed" travel.

A critical point regarding model validation is that every component of a model must be validated, as well as the entire model system. For conventional four-step travel models, this includes the four major components – trip generation, trip distribution, mode choice, and mode specific trip assignment – along with the model input data and other components that might be part of the model system, such as vehicle availability, or time-of-day modeling processes. For a more complex model system, such as an activity-based modeling, there could be 10 to 20 model components requiring validation.

Validation Sequence

The recommendation to perform validation of both model components and the entire model system may result in a validation sequence conundrum. Specifically, a determination must be made regarding the amounts of time and resources to devote to component validation as the components are developed versus waiting until all components have been estimated and (initially) calibrated to perform component validation in conjunction with system validation. Both approaches can have benefits and costs. A simple example of the conundrum can be illustrated with trip generation models. Trip generation can be validated against calculated trip rates by comparing to expanded household survey data, but issues with underrepresentation of total trips by the household survey may not be realized until the system validation is performed and modeled versus observed vehicle-miles of travel comparisons are performed.

Validation sequence should be considered in the validation plan specification. It is impossible to complete the model system validation prior to the completion of the model component validations. Establishing a sequence, where a model component is initially validated and applied for an initial full system "validation" using existing model components for the subsequent steps, might be an efficient approach. It is inevitable that some iteration will be required in the validation sequence for model components and the overall model system.

Types of Validation Checks

Four broad categories of validation checks will be used in this manual:

1. **Comparisons of base year model results to observations** might be considered "traditional" validation. The comparisons might be of model results to disaggregate data such as data from a supplementary survey not used for model estimation or to aggregate data such as traffic counts or transit boardings. Comparing base year model results to different aggregations of the data used to estimate or calibrate a model is not as sound of a validation practice as comparing to independent data. However, for some validation tests, the data used for model estimation or calibration are the only data available.

2. **Temporal validation** is an important aspect of model validation since, by definition, it implies comparing model results to data not used in model estimation. Both backcasts and forecasts may be used for model validation. For example, if a model is estimated using 2007 survey data, the model could be used to backcast to 2000 conditions, and compared to year 2000 traffic counts, transit boardings, Census Transportation Planning Package (CTPP) data, or other historical data. Likewise, if a model was estimated or calibrated using 2005 survey data, a "forecast" validation could be performed against 2008 data.

3. **Model sensitivity testing** includes several important types of checks including both disaggregate and aggregate checks. Disaggregate checks, such as the determination of model elasticities, are performed during model estimation. Aggregate sensitivity testing results from temporal validation. Sensitivity testing can also include model application using alternative demographic, socioeconomic, transportation supply, or policy assumptions to determine the reasonableness of the resulting travel forecasts.

4. **Reasonableness and logic checks** include the types of checks that might be made under model sensitivity testing. These checks also include the comparison of estimated (or calibrated) model parameters against those estimated in other regions with similar models. Reasonableness and logic checks may also include "components of change" analyses and an evaluation of whether or not the models "tell a coherent story" as recommended by the FTA for New Starts analysis.

■ 1.5 Validation Considerations

1.5.1 Types of Travel Demand Forecast Model Applications

When travel demand forecasting models were first developed, the transportation planning questions that they were used to address were very different questions from today. The initial use of travel forecasts focused on highway design and location. The outputs from the models were typically used for elements such as roadway sizing, geometric design, and pavement design. Over time, the role of the model and forecasts have evolved to cover not only roadway design concerns, but also transit system planning, air quality issues in support of conformity analysis, evaluation of travel demand management (TDM) programs, highway noise abatement programs, as well as social policy and environmental justice.

While the demands on models may change, models are originally developed with the goal of providing information for specified types of analyses. The types of analyses to be addressed, the scope (or area covered) and level of detail for the anticipated analyses, and the types of information needed from the analyses should be considered when performing model validations.

Validation for the "Unknown"

Model validation must be driven by the intended uses of the models. Yet, travel forecasts are being used to provide input for the analysis of numerous and diverse transportation, environmental, financial, social, and land use issues that may not exist when the models are developed. For example, a model calibrated to adequately represent current ridership on a bus system carrying predominantly captive, walk-to-transit riders may be called upon to forecast choice responses to a significantly different transit system with park-and-ride facilities and/or "fast" transit service. Traveler response to congestion pricing is another example where many models may be called upon to forecast the unknown.

It is impossible to anticipate all of the questions that the models may be called upon to help address. When new questions are asked, model validation should become an ongoing effort to determine the appropriateness and sensitivities of models. The ongoing validations may require the collection of new data, may lead to revisions to the model, or may result in the conclusion that the travel model is not sensitive to the issue being considered.

Role of Temporal Validation

Most travel models are based on "snapshot" data, such as household survey data collected in a periodic, but infrequent, survey effort. The model relationships, parameters, and coefficients might be significant and accurately reproduce travel for the point in time represented by the model estimation data. However, the relationships may not hold true over time; the further one moves from the base year for validation, the more uncertain one should be regarding the veracity of the models. For this reason, good validation practice should include temporal validation for at least one year other than the base year for model estimation or calibration. The temporal validation should be for a year for which some validation data, such as traffic counts or transit boardings, are available.

1.5.2 Accuracy Requirements and Guidelines

Accuracy requirements and guidelines for model validation depend on the intended use of the model being validated. Models used for project design or comparing alternative projects might require tight matches between modeled and observed travel data for model validation. In other cases, such as the evaluation of alternative transportation policies, the correct sensitivity of the model might outweigh the need for a close match of observed data. The varying uses and requirements of travel models has led some MPOs to develop advanced modeling techniques such as activity-based or tour-based models in an effort to respond to a wider range of questions. Alternatively, the varying uses and requirements of forecasts could lead to the development of multiple models for a region or multiple application approaches for a single model.

The “close enough” point of view outlined above must be weighed against economic realities affecting many state Departments of Transportation (DOTs) and MPOs. Most users of the models and forecasts would like models that can respond to all issues and transportation options. Most DOTs and MPOs develop a single model for an area (referred to in this manual as the “modeled region”) and use it to provide base forecasts for all analyses. The desire to use a single model might become even more prevalent as increasing infrastructure needs coupled with decreasing revenues result in shrinking modeling budgets. This calls for better guidance regarding good modeling and validation practices. Claiming acceptability for a model that fails to achieve specified values for metrics such as percent root mean square error, screenline crossings, and vehicle miles traveled ratios might seem irrational to a decision-maker if other agencies not using acceptable modeling and validation procedures publish better “validation” results based on a few selected metrics.

Role of Reasonable Validation Thresholds

Reasonable validation thresholds may be important in helping establish the credibility of a model and helping model developers and users determine when the model is “close enough.” The definition of an acceptable threshold should be a local decision and needs to balance the resources and time available for model development with the decisions that will be supported by the travel forecast obtained using the model.

The term “threshold” rather than “standard” will generally be used throughout this manual. The term standard connotes a formal definition of acceptance: “The standard has been met, therefore the model is valid.” While it is important to match base year observations for validation, simple matching of traffic counts, for instance, is not sufficient to establish the validity of a travel model. Quality model validation must test all steps of the travel model and also should test model sensitivity. If standards are set for models by agencies or model reviewers, it is beneficial that they not convey a formal definition of acceptance but, rather to help set boundaries or levels of confidence regarding the use of travel forecasts for studies.

Accuracy Requirements for Different Forecast Uses

There are different levels of transportation planning studies ranging from the simple traffic impact analyses to complex regional policy and planning studies. Each type of study may address a different timeframe, require different modeling tools, and be subject to different accuracy requirements. Accuracy requirements will generally decrease as the planning horizon moves increasingly into the future, as anticipate financial outlays decrease, and as planning issues become more ambiguous. As outlined above and covered in more detail in Chapter 2, Model Validation Plan Specification, the uses of the model should be understood when developing and validating travel models.

Acceptable Methods for Achieving Validation Thresholds

The acceptability of methods for achieving validation thresholds will depend, in part, on the type of questions being addressed using the travel models. For example, NCHRP 255, Highway Traffic Data for Urbanized Area Project Planning and Design, describes post-model factoring procedures that can be used to adjust traffic forecasts. The procedures have been used frequently and have helped improve traffic forecasts for project planning and design. The techniques, however, are applied for a specific planning context and are not generally acceptable for all planning studies.

In general, the following guidelines should be used to determine acceptable methods for achieving improved match between modeled and observed travel characteristics:

- The adjustments should reflect transportation supply or traveler behavior rather than simple arithmetic;
- The adjustments should be reproducible; and
- The reasons for adjustments should be clearly documented.

1.5.3 Level of Aggregation

The level of aggregation for model validation is an important consideration from two different standpoints. The first, obvious, issue of aggregation level relates to geographic aggregation. A model might be validated on a regional basis using regional criteria. However, for a subarea or corridor study, the regional validation criteria might be insufficient to demonstrate the veracity of the model for the smaller area.

The second aggregation issue relates to the validation tests and data. There is a continuum of checks ranging from validation using disaggregate data at the traveler or household level to aggregate results at the regional level. In the middle are validation checks using the models applied to zonal data. For state-of-the-art disaggregate models, the entire range of checks is needed to ensure that the models can reproduce not only the travel behavior of individual households, but also the resulting performance of the transportation system when all of the individual trips are aggregated over the entire modeled region. The two ends of the continuum are:

- **Disaggregate validation** provides a means of exploring how well a candidate model fits the observed data at the household or individual level. It involves defining subgroups of observations, based, for example, on household size and income or auto ownership levels. Model predictions are compared with observed data to reveal systematic biases.

- **Aggregate validation** provides a general overview of model performance through regional travel characteristics such as average trip rates, average trip lengths, average mode shares, and regional vehicle miles traveled (VMT). Travel models have traditionally been applied to aggregate data at the regional, county, district, or zonal level. Emerging travel modeling procedures may include population synthesizing techniques allowing travel models to be applied at the household or person level. Traffic assignment results are validated at a regional level, using screenline volumes, and then at a local level, using cutline and individual link volumes.

1.5.4 Sources of Error

There are a number of sources of error impacting travel models, including:

- **Model specification error** is the error introduced by imperfect understanding or accounting for traveler behavior, traveler response to transportation system changes, and transportation supply.

- **Model aggregation error** is the error introduced by the abstraction and aggregation inherent in travel models. Due to limitations in understanding and computer processing, models are aggregated over travelers (e.g., travelers may be aggregated over households or zones), traveler decisions (e.g., trip purposes), time periods, and transportation networks.

- **Model estimation data error** is the error introduced through the collection and processing of data for model estimation. The error may be random error associated with the collection of survey data or it may be introduced through improper data collection, checking, or editing. Error introduced through improper data collection, checking, and editing may be correctable based on model validation findings and used to improve the model estimation or calibration.

- **Input data error** includes mistakes in input data used by the models. This type of error may be reduced through the model validation process through the correction of transportation networks and network processing procedures, or through the correction of population and employment estimates.

- **Validation data error** is similar to model estimation data error. As with model estimation data error introduced through improper data collection, checking, and editing, validation data error may be corrected based on model validation findings. However, correction of validation data error will not require a return to the model estimation or calibration.

A major concern for validation of travel models is error inherent in the collection of input data or historical data used for validation. Problems with input data or validation data can lead to erroneous corrections to models that, ultimately, will damage model performance, credibility, and results. For example, if daily traffic counts collected at screenlines are low due to incorrect collection methods, an analyst may attempt to increase auto occupancy rates or lower trip rates in order to match the screenlines. Thus, a good course of action for models that do not validate satisfactorily is to check for errors in model input or validation data prior to returning to the model estimation or calibration steps.

1.5.5 Validation Responsibilities

Responsibilities for model validation will vary from state to state and region to region. The following outlines typical or traditional responsibilities for model validation.

Agency Developing/Maintaining the Model

The responsibly for the model validation typically falls on the agency that develops and maintains the model. In many regions, this responsibility falls on the MPO for the region or on the state DOT.

Agencies Supporting/Using the Model

Many regional models are used by agencies that do not develop the models. For example, a transit agency may use the regional travel model developed and maintained by the MPO. Depending on the relationships of the agencies, model validation for directed uses of the model (e.g., transit ridership forecasts) may fall upon the agency using the model or the agency developing the model. Funding for additional model validation required by agencies using the model might be provided to the agency developing and maintaining the model for the region.

Model Developer

Frequently, a model might be developed for a region or agency by a consultant. Model validation responsibilities may be assigned to the model developer, to the agency sponsoring the model, or to another entity. Regardless of who is responsible for model validation, it is a good idea to include involvement by the model developer due to the developer's knowledge of the model estimation and calibration data and techniques.

Consultants

Consultants frequently use travel models to support private or public clients. While model validation responsibilities will be dictated by contractual relationships, consultants should be familiar with good validation practices in order to advise their clients on proper validation efforts required for a study.

Federal and State Agencies

Federal and state agencies often provide direction for model validation efforts. Direction may be provided directly through the development of validation manuals and guidelines such as this document or indirectly through approval of travel forecasts resulting from validated travel models.

1.6 Organization of Manual

The remainder of this manual is divided into the following chapters:

- **Chapter 2** discusses the development of a model validation plan to guide the model validation process and associated validation data collection tasks.

- **Chapter 3** discusses the validation of model inputs and reasonableness checks for input data, including land use and socioeconomic data and transportation networks.

- **Chapters 4 through 9** discuss validation techniques and reasonableness checks for model parameters and outputs for individual model components. The individual chapters generally address components of travel following the trip-based travel modeling process. However, the chapters have been augmented to address emerging activity- and tour-based travel modeling techniques. Individual chapters address:

  - Chapter 4 – Socioeconomic Models;
  - Chapter 5 – "Amount of Travel/Activity" (Activity Patterns, Trip/Tour Generation);
  - Chapter 6 – Trip Distribution/Destination Choice/Location Choice;

- Chapter 7 – Mode Choice/Auto Occupancy;
- Chapter 8 – Time-of-Day Choice/Time-of-Day Split/Directional Split Factors; and
- Chapter 9 – Assignment Procedures.

- **Chapter 10** discusses the roles of temporal validation and sensitivity testing in the model validation process.
- **Chapter 11** discusses the importance of model validation documentation in the overall model validation process. This chapter includes suggestions regarding information to include in model validation documentation.

■ 1.7 Acknowledgments

This edition updates the first edition published in 1997[4]. The current edition was prepared by Cambridge Systematics under contract to FHWA. The work was performed under a task order administered by the Texas Transportation Institute (TTI), as part of its contract to provide outreach services for FHWA's Travel Model Improvement Program (TMIP).

Thomas Rossi and David Kurth of Cambridge Systematics were the principal authors of this report. Sarah Sun of FHWA oversaw the report's development, and Gary Thomas of TTI oversaw the task order.

A panel of travel demand modeling practice experts reviewed the report and provided valuable comments to improve it. These experts included Paul Agnello of the Virginia Department of Transportation, Ken Cervenka of the Federal Transit Administration, Alan Horowitz of the University of Wisconsin, Milwaukee, Ken Kaltenbach of The Corradino Group, Guy Rousseau of the Atlanta Regional Commission, David Schmitt of AECOM, and Ed Weiner.

The authors express their sincere thanks to all of those who contributed their time and expertise to this effort.

---

[4] Model Validation and Reasonableness Checking Manual, prepared by Barton-Aschman Associates, Inc. and Cambridge Systematics, Inc. for Federal Highway Administration Travel Model Improvement Program, February. 1997

2.0 Model Validation Plan Specification

2.0 Model Validation Plan Specification

You've got to be very careful if you don't know where you are going, because you might not get there.[5]

2.1 Validation Plan Specification

Most travel model development and update programs follow carefully planned approaches. For example, initial decisions are made regarding likely model forms; the availability and quality of existing model estimation data are evaluated; decisions are made regarding the need to collect additional data for model development; a model development plan is established (possibly as a request for proposals from consultants); and model estimation and calibration are initiated. The specification of such plans enhances the likelihood of a successful model development process and provides a basis for budgeting for the model development process.

Likewise, the development of a model validation plan will enhance the likelihood of a successful validation process. A successful validation process will, in turn, lead to greater acceptance of travel forecasts and, hopefully, improved decision-making regarding the expenditure of scarce public funds on transportation projects.

2.1.1 Timing of Validation Plan Specification

Model validation plans are best specified at the outset of a model development process when important decisions are made regarding:

- The availability of validation data and the need to collect additional data;
- The level of validation required based on likely uses of the model;
- Proper goals for the accuracy and sensitivity of the model and model components considering the proposed uses for model results;

---

[5] Yogi Berra, BrainyQuote.com, Xplore Inc, 2010, http://www.brainyquote.com/quotes/quotes/y/yogiberra124868.html, accessed April 21, 2010.

- Tradeoffs regarding the allocation of funds for model development and model validation; and

- The proper allocation of funds for validation data collection, validation efforts, and sensitivity testing.

Even if a model validation plan is not specified at the outset of the model development process, it is useful to specify a plan prior to the initiation of model validation in order to establish guidelines for the accuracy and sensitivity of the model and model components that are consistent with the proposed use of model results.

2.1.2 Recommended Validation Plan Components

A model validation plan should include the following components:

- Identification of validation tests,
- Identification of validation data, and
- Identification of validation costs

Each of these validation plan components will be covered in additional detail below. In addition, subsequent chapters of this manual discuss various validation tests that can be considered in the validation plan.

2.2 Identification of Validation Tests

2.2.1 Validation Context

Model validation must be driven by the intended use of the models, including the **types of analysis** being supported by the model, the **scopes of the analyses**, and the **business processes** being supported by the model. Validation tests and standards might be different for each intended use. Thus, a single travel model might be subject to multiple validation efforts.

Types of analysis may be as divergent as policy analysis and project planning. The model validation required for planning a highway expansion will differ from model validation for policy analysis. Two basic modeling needs, highway system planning and New Starts applications, require model results to satisfy more rigorous standards regarding their ability to match traffic counts or boarding counts. In this context, the focus of model validation must be whether the model is representing reality. Proper model sensitivities are important for both project planning and policy testing. However, the need to reproduce observed roadway and transit volumes might be less crucial for models used for policy testing if the reduced accuracy results in increased, but reasonable, model sensitivities.

Policy questions may appear suddenly and often lead to questions regarding model sensitivity. Often the model is expected to respond adequately to policy questions that were not considered when the model was developed. In such cases, model validations demonstrating appropriate sensitivities in response to different scenarios are important. For example, the sudden increase in fuel prices in the summer of 2008 led to validation concerns regarding the use of fuel prices in models as well as the sensitivity of models to changes in fuel price.

Scopes of analysis may include such uses as regional planning, systemwide planning, corridor planning, interchange justification reports, or site impact analyses. Systemwide planning- and project-level planning require different validation methods. For systemwide planning, validation should combine model sensitivity testing with testing based on matching absolute aggregate numbers, such as independent estimates of VMT. For project-level planning, a focus on matching more detailed absolute numbers might be most appropriate. For example, matching time-of-day traffic volumes and traffic speeds might be important validation measures for project-level planning.

The planning horizon is also an important scope of analysis consideration. Typically, project-level analyses are assigned short-term planning horizons while the systemwide analyses have long-term planning horizons. For the short term, criteria based on more detailed absolute numbers, such as time-of-day traffic volumes, become more important in validation efforts, since more detailed project decisions may be made based on the forecasts. Long-term planning usually focuses on more general goals and objectives regarding regions and the development of programs and allocation of resources to achieve those goals and objectives. Thus, validation should focus on model sensitivity to factors impacting travel decisions and traveler behavior.

Temporal validations such as forecasting or backcasting are important for systemwide model validations, particularly if sufficient time or transportation system changes exist between the years selected for the validation. While sensitivity testing is most often considered for long-term planning, it is also an appropriate validation test for short-term planning.

A full range of types of "absolute numbers" may be considered as validation of travel models moves from long-term regional planning to short-term project-level planning. For the long-term regional planning context, the absolute numbers considered include regional VMT, regional transit boardings, regional mode shares, and major screenline crossings or district-to-district flows. Model validations for short-term corridor planning, interchange justification reports, and site impact analyses focus on reproducing link specific traffic counts or detailed transit line boardings.

Business processes may include the planning efforts required to support an MPO, a regional transit district, or an air quality control district. The business process relates to the charge of the agency using the travel forecasts. A primary objective for model validation is an improved tool to support decision-making. This requires that decision-makers and other users are satisfied that the model structure and results address their needs. Considering the business process provides a practical context to model validation.

2.2.2 Types of Validation Checks

There are three basic types of validation checks to consider in a model validation plan: disaggregate checks, aggregate checks, and sensitivity/reasonableness checks. While each may be used to support the various validation contexts discussed above, some might be more appropriate in certain contexts.

Disaggregate checks are characterized by the use of disaggregate data. These checks include the determination of elasticities for various model coefficients and the comparison of modeled to observed travel choices for individual trips, travelers, or households. Disaggregate checks may not be appropriate for all model components. For example, it is not feasible to perform disaggregate checks of traffic assignment for a static equilibrium assignment process. However, if geographic positioning system (GPS) units are used for travel survey data collection, it might be possible to validate a traffic simulation procedure.

Ideally, validation should be performed using data other than those used for model estimation. As a result, disaggregate checks may be more difficult and costly to perform than other types of checks. For instance, for an ideal validation, a travel survey dataset used for model development must be sufficiently large to divide into model estimation and validation datasets or two travel survey efforts might be required to collect model estimation data and model validation data.

The collection of two survey datasets might be a feasible approach over time. For example, the Puget Sound Regional Council (PSRC) travel model was estimated in 2000 based on 1999 household survey data. The PSRC collected a new household travel survey (more than 4,000 households) in 2006, providing them the option of performing a disaggregate model validation of the year 2000 travel model using 2006 travel survey data.

Disaggregate checks focus on the reproduction of traveler behavior, and are therefore more appropriate for validation contexts that test model sensitivities.

Aggregate checks are characterized by the use of aggregate data. These checks are the most frequently used since aggregate validation data are most commonly available. Aggregate checks can include comparing how closely the travel model reproduces traffic counts and transit boardings, regional VMT, mode shares, and district-to-district trip flows. Aggregate checks are more focused on the reproduction of travel patterns and are therefore more appropriate for validation contexts that require good traffic and transit forecasts.

Sensitivity/reasonableness checks are typically focused on the response of models to changes in transportation supply or policy. Sensitivity checks may be considered validation checks if they are based on forecasting or backcasting of travel, especially if there has been a major change in transportation supply or policy. For example, the Denver Regional Council of Governments (DRCOG) travel model was estimated and calibrated using 1997 travel survey data and validated against 2005 traffic counts and transit boarding data. A major extension of the Southwest light-rail transit (LRT) line was performed between 1997 and 2005, providing the opportunity to test the sensitivity of the model in a “real world” setting. In rare cases, such as with the 1999 and 2006 PSRC data, disaggregate sensitivity checks that are true validation checks may be performed.

Reasonableness checks focus on the rationality of travel model response to transportation supply or policy changes. Reasonableness checks may be aggregate or disaggregate, but are not true validation checks since they are not compared to observed travel data.

Risk analysis is the evaluation of impacts that may result from inaccurate forecasts. Risk analysis is not a model validation practice but, rather, a travel model forecast consideration associated with specific alternatives being evaluated. Risk in travel forecasts for projects is introduced by many sources, only one of which is the travel model. Nevertheless, model validation may contribute information for risk analyses by providing “confidence intervals” for the model based on the different types of validation checks defined above. The specification of the model validation plan might include the collection of validation data and identification of validation checks designed to provide useful information for future risk analyses.

2.2.3 Specifying Validation Expectations

A model validation plan should include a listing of the model components to be validated, the validation tests to be performed, the aggregation level for those tests, and the expected outcomes. If required, standards may also be set. As examples, an excerpt from the DRCOG Integrated Regional Model Validation Plan pertaining to one model component (the daily activity pattern model) is shown in Figure 2.1, and an excerpt from the *FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards: Model Validation Guidelines and Standards* is shown in Figure 2.2.

Figure 2.1 Excerpt Detailing Planned Validation Tests from DRCOG IRM Validation Plan

Table 3.6 Daily Activity Pattern Model Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Comparison of model parameters to other regions	• Comparison of model coefficients to: • Sacramento • San Francisco • Columbus	• No expectations; comparison only.	Level 1
Disaggregate	• Prediction success of modeled daily activity pattern choices against observed TBI estimation data	• Prediction success likely to be very low	Level 3
Aggregate	• Numbers or percents of residents making tours and intermediate stops by activity type: • For the region • By county • By household size and income group • By household size and auto ownership • By gender and age group • By employment status • By student status	• Compare modeled to expanded observed numbers or percents • Review for reasonable patterns	Level 2
	• Percent of "immobiles" (persons with no out of home activities during the day) by: • By household size and income group • By household size and auto ownership • By gender and age group • By employment status • By student status	• Compare to results summarized by Kay Axhausen (e.g., in *Transportation*, Volume 34, Number 1, January 2007, pp. 107-128)	

Source: Parsons Transportation Group, *DRCOG IRM Validation Plan Technical Memorandum*, Draft 2a, September 2007.

Figure 2.2 Excerpt Detailing Planned Validation Standards from FSUTMS Validation Guidelines

Table 3.9 Volume-Over-Count Ratios and Percent Error

Standards	Statistic Acceptable	Preferable
Freeway Volume-over-Count (FT1x, FT8x, FT9x)	+/- 7%	+/- 6%
Divided Arterial Volume-over-Count (FT2x)	+/- 15%	+/- 10%
Undivided Arterial Volume-over-Count (FT3x)	+/- 15%	+/- 10%
Collector Volume-over-Count (FT4x)	+/- 25%	+/- 20%
One way/Frontage Road Volume-over-Count (FT6x)	+/- 25%	+/- 20%
Freeway Peak Volume-over-Count	75% of links @ +/-20%	50% of links @ +/-10%
Major Arterial Peak Volume-over-Count	75% of links @ +/-30%	50% of links @ +/-15%
Assigned VMT-over-Count Areawide	+/-5%	+/-2%
Assigned VHT-over-Count Areawide	+/-5%	+/-2%
Assigned VMT-over-Count by FT/AT/NL	+/- 25%	+/- 15%
Assigned VHT-over-Count by FT/AT/NL	+/- 25%	+/- 15%

Source: *FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards: Model Validation Guidelines and Standards*, prepared for Florida Department of Transportation, prepared by Cambridge Systematics, Inc., December 31, 2007.

2.3 Identification of Validation Data

A model validation plan should include a listing and assessment of the data available for validation. Appropriate validation tests can then be defined based on the assessment of the available data. The listing and assessment also provide direction for the collection of addition validation data. An example of an assessment performed by the Sacramento Area Council of Governments (SACOG) is provided in Figure 2.3.

Figure 2.3 Draft Validation Data Sources Assessment from SACOG

Portion of Entire Assessment Shown

Monitoring Dataset	Comprehensive-ness	Currency	Consistency	Content	Geographic Scale	Travel Model Validation for…
Locally-Developed GIS						
SACOG Centerline GIS	Regional	Current (Annual Updates)	Well Defined Data Standards	Linear feature alignment; street addresses (by range); limited information on type of feature	Micro	Roadway alignment; potentially roadway distances; walk distances
Census Data						
Year 2000 Short-Form (STF1)	Regional	8 years old, decenial updates	Generally comparable, one decade to next. Thick sample (1:1)	Detailed aggregate demographics, cross-tabs	Very small area (census block)	Aggregate-by-area cross checks on population files and zonal datasets (pers/hh, workers/hh, hh income, auto ownership, etc.)
Year 2000 Long-Form (STF3, CTPP)	Regional	8 years old, decenial updates	Generally comparable, one decade to next. 1:5 sample	Detailed aggregate demographics w/cross tabs; journey-to-work travel data	Small area (block group)	Worker flows; home-workplace distance distribution
ACS	Currently, census places >65k in pop; by 2010, >20k	Rolling 3-year sample	Thin sample each year: 1:20	Detailed aggregate demographics w/cross tabs; journey-to-work travel data	???	TBA, once reported geography gets below places 65k and greater (2009 or 2010)
Travel Surveys						
SACOG Household Travel Survey	Regional	8 years old; 2010 update planned	Not comparable to 1991 survey; very thin sample (1:250)	Detailed disaggregate (person level) data; includes detailed demographic and trip-level information on all purposes (including non-work)	Micro (parcel/point)	Once expanded, limited/weak checks of tour frequency, home-to-tour-destination distance distribution, mode of travel by purpose and person type, etc.
2005 Transit On-Board Survey	All fixed-route operators	3 years old; no update planned	Not comparable to 1999 survey; 1:10 sample	Detailed disaggregate (passenger trip-segment level) data; includes some demographic and whole trip-level information.	Varies; mostly parcel/point locations	Aggregate checks on tour and trip mode choice; aggregate checks on transit assignment (boardings/trip, etc)
National Household Travel Survey	National	2001, update ongoing	Evolving survey instrument, but largely comparable for trend analysis	Reported detailed aggregate travel behaviour, cross-tabulated by demographics, area of residence, etc.	Some states have special add-ons; otherwise, national	Reasonable-ness checks on basic travel behavior (e.g. trips per person, per hh; VMT per person, per hh; etc.
Transportation Network/Supply						
Nat'l Transit Database	Selected fixed route operators	2006, w/annual updates	Generally comparable to prior years	Systemwide supply (revenue miles, revenue hours) by bus vs. rail; weekday vs. weekend values	Operator totals	Aggregate transit network stats by operator
HPMS	Regional	2006 available; annual updates	Generally comparable to prior years	Aggregate network supply (lane miles, centerline miles)	Jurisdiction, unincorporated remainder by county	For freeways, comparable to model network; includes all local streets (not included in model network), so lower level capacity classes not comparable to model.
Transportation System Utilization						
Traffic Counts	Regional, but very spotty locations	2005; 2008 ongoing	Comparable to some locations in 2000	Varies: all include typical weekday totals; most include hourly volumes by direction; quality varies, too: some counts are robust averages, some are single counts	Very spotty (e.g. about 1,000 locations, with some jurisdiction uncounted)	Aggregate traffic assignment validation, by: functional class of roadway, time period, link volume group
Transit Line Counts	All fixed-route operators	2005; 2008 ongoing	Comparable to 2000 data	Weekday averages for all in 2005	n/a	Aggregate transit assignment validation: daily boardings by line, operator, service type.
LRT Station Boardings	All LRT Stations	2005; 2008 ongoing	Comparable to 2000 data	Spring/fall weekday averages for 2005, by RT service period	By station	Aggregate station boardings, daily and by time period.
Park-and-Ride Lot Occupancy	All LRT lots; some other lots	2005; 2008 ongoing	Available back to ???	RT collects monthly; SACOG uses spring/fall	By station	Peak park-and-ride demand by station
HPMS	Regional	2006 available; annual updates	Generally comparable to prior years; questions as to the frequency of local jurisdiction counts/volumes	VMT aggregated to roadway class	Jurisdiction, unincorporated remainder by county	VMT by county
Nat'l Transit Database	Selected fixed route operators	2006, w/annual updates	Generally comparable to prior years	Systemwide ridership (boardings, passenger miles, passenger hours) by bus vs. rail; weekday vs. weekend values	Operator totals	Aggregate ridership statistics by operator

2.3.1 Assessing Currently Available Validation Data

How well the validation data represent reality is a primary validation question. This question can be illustrated by a review of the veracity of commonly used validation data, traffic counts. Counts are often collected from multiple sources using multiple counting techniques. They may be stored as raw counts or factored counts, such as average annual daily traffic (AADT). Developing a validation dataset of average weekday traffic (AWDT) may be difficult due to the different sources, different counting methods (one-day, two-day, permanent traffic recorder), and reporting methods (raw axle counts, raw counts divided by average axle factors, AADT estimated from raw counts).

Even when traffic counts are collected and stored in a consistent manner, there can be substantial variation in day-to-day counts. Figure 2.4 summarizes data collected in 1994 from 21 continuous count stations in Florida. The number of count days at the sites ranged from 210 to 353. The counts were used to estimate the average annual daily traffic (AADT) for each site along with the standard deviations around those means. Figure 2.4 shows an "error bar" representing ±1.96 standard deviations as a percent of the AADT for each of the count sites. While it is not precisely correct in terms of statistics, roughly 95 percent of the daily counts should be expected to be within ±1.96 standard deviations of the AADT.

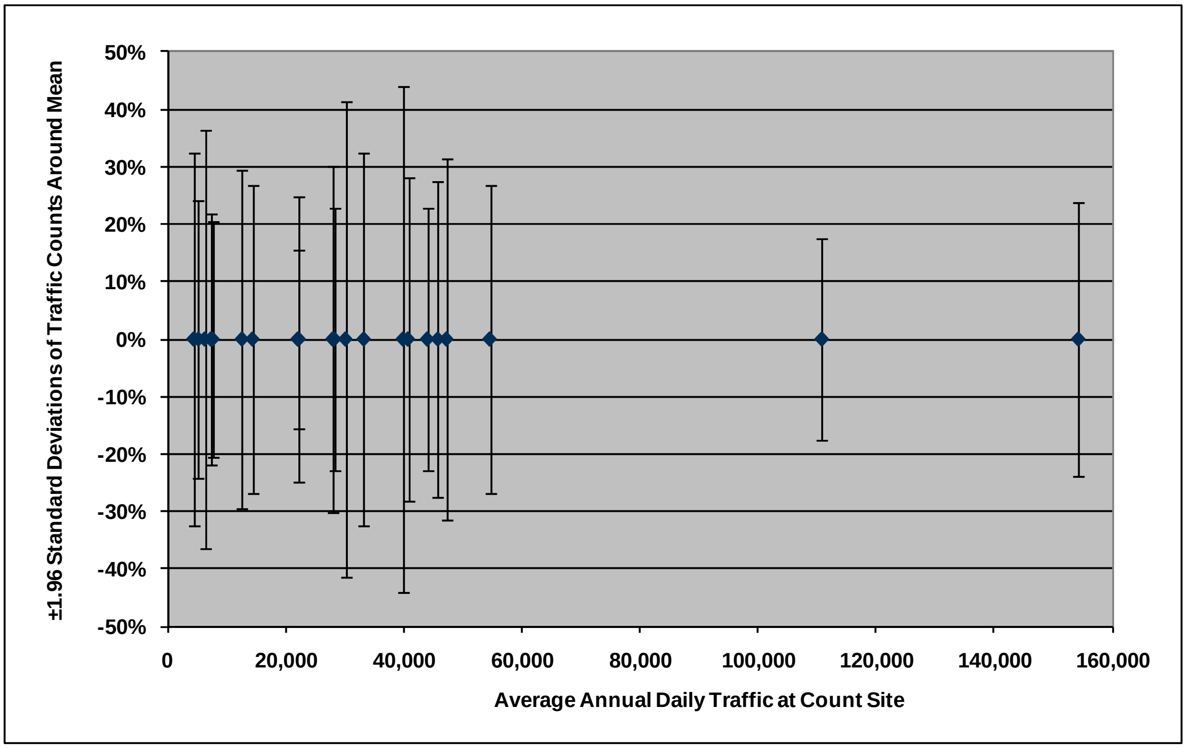
While the standard deviations shown in Figure 2.4 should have been less if only weekday traffic had been considered, the analysis underscores the issue of variability associated with all observed data collected using sampling procedures. Unfortunately, it is not always obvious how data have been collected or how much sampling error is inherent in the data.

Similar issues and concerns can be raised with many other types of data used for model validation purposes. Thus, as shown in Figure 2.3, an assessment of the data quality as well as the data availability should be performed for the development of any model validation plan.

2.3.2 Prioritizing Validation Data Collection Needs

The assessment of validation data coupled with the development of the checks to be included in the model validation plan can guide the setting of priorities for validation data collection. Specifically, if validation test priorities are established in the validation plan (see Figure 2.1), the data required for the highest-priority tests can be reviewed for availability and quality. Decisions can then be made regarding which data collection efforts will be most cost effective for improving overall model validation.

Figure 2.4 Variability in Daily Traffic Count Data for 21 Sites in Florida



Source: Wright, Tommy, et al., *Variability in Traffic Monitoring Data, Final Summary Report*, prepared for Oak Ridge National Laboratory, August 1997, Table 5, Page 10.

Proper collection of validation data is as crucial as proper data collection for model estimation and calibration. Improper validation data can lead to poor decisions regarding the veracity of the travel models whereas well thought out and properly designed data collection efforts can enhance overall model validation. For example, a transit on-board survey collection effort might be necessary for model validation in preparation for a New Starts application. In addition to the development of estimates of linked transit trips for mode choice model validation, the data may be used to validate transit path-building procedures and district-to-district transit flows if the data collection effort is properly designed and conducted. In this example, it might be possible to validate the two additional model components (transit path-building and transit trip distribution) for relatively little additional cost with the result being a better overall model validation.

2.4 Validation Budgeting

In the May 9, 2008 Peer Exchange on Travel Model Validation Practices, participants were asked for their recommendations on allocation of resources for data collection and model estimation, calibration, validation, and documentation. The results are summarized in Table 2.1. Some care should be used in interpreting the recommendations shown in the table. Peer exchange panelists based their recommendations on the definitions of model estimation, calibration, and validation shown in Chapter 1, which may vary somewhat from definitions used in specific regions. Quite often, model calibration is mixed with model estimation or with model validation. The information summarized in Table 2.1 suggests that these three efforts (estimation, calibration, and validation) should comprise approximately 50 percent of a model development budget and that budget allocations for each of the three should be approximately equal.

As shown in Table 2.1, almost 40 percent of an overall budget for model development is recommended for data collection. The data collection effort should include data for both model estimation and model validation. The overall allocation between those two broad needs will vary by region depending on assessments of local needs and priorities, currently available model estimation and calibration data, and currently available validation data.

Table 2.1 Informal Survey of May 7, 2008 Travel Model Validation Practices Peer Exchange Participants Regarding Allocation of Modeling Resources

e	Desired Resource Allocation (Percent)[a]
Data collection	39
Estimation	16
Calibration	17
Validation	17
Documentation	9

[a] Percents do not sum to 100 due to rounding.

3.0 Validating Model Inputs

3.0 Validating Model Inputs

There are two broad types of data required for any travel- or activity-based modeling process: socioeconomic data and transportation network data. These two broad types of data are the basic building blocks used along with the travel models to estimate or forecast travel in a region, ultimately expressed as traffic on roadways or riders on transit services. Good base year socioeconomic and network data impact model estimation, model calibration, and model validation. Reasonable future year travel forecasts require reasonable future year socioeconomic and network data forecasts. Thus, the success or failure of the modeling process rests on the input data. The old adage "garbage in, garbage out" is appropriate.

The term "socioeconomic data" is used in this chapter to represent the bases for generating the demand for travel or activity generation. In this chapter, socioeconomic data may be interpreted to include inputs such as population, households, group quarters population, and employment. For land use-based models, the term may be expanded to include the area or gross square feet of different land uses or counts such as the number of dwelling units or number of seats in an entertainment venue. In some models, the basic input data may be stratified by different characteristics; households may be stratified by income group or employment stratified by employment type.

Socioeconomic data are typically compiled and coded to transportation analysis zones (TAZs, sometimes called traffic analysis zones). In the past, it has not been feasible to represent every household, place of employment, shopping center, and other activity as a separate point so the data were aggregated into TAZs. Some modern models are beginning to use parcel-based land use data. Regardless of how the data are stored in the travel models, procedures should be in place to aggregate the data to larger geographic units such as districts or into political units such as cities or counties. Table 3.1 provides an example of socioeconomic data used in a typical travel model using TAZs. Many models will have more or different data items than those shown in this table.

Transportation network data are the basic descriptors of the available transportation supply. Networks include roadway (often referred to as highway networks), transit, and in some emerging model systems, nonmotorized mode networks. Roadway networks may include representations of roadways designated for general purpose, single-occupant vehicle (SOV), high-occupancy vehicle (HOV), and/or truck use. These uses are usually represented in one composite transportation network through a special link type variable. Then, for trip distribution and mode choice, mode specific paths and skims are created by using the appropriate network specified using link types. For example, HOVs can travel on HOV links in addition to all SOV links. An example of roadway network data used in a typical travel model is shown in Table 3.2. Many model networks will have more data items or different data items than those shown in this table.

Table 3.1 Example Socioeconomic Data for Travel Models

Data Item	Description/Use
TAZ	Transportation Analysis Zone Identifier
DISTRICT	District designation for summaries
JURISDICTION	Political jurisdiction for summaries
AREA	TAZ
HH_POP	Total household population (excluding group quarters population)
MEDINC	Median household income (1999 dollars)
HHOLDS	Total households
LOWINC_HH	Number of low income households (bottom 11%)
MEDINC_HH	Number of medium income households (middle 64%)
HIGHINC_HH	Number of high income households (top 25%)
PROD/DIST_EMP	Total production and distribution employment (NAICS Codes 11-42, 48-49)
RETAIL EMP	Total retail employment (NAICS Codes 44-45)
SERVICE EMP	Total service employment (NAICS Codes 51-92)

Table 3.2 Example Input Roadway Network Data for Travel Models

Data Item	Description/Use
ANODE	From node of link
BNODE	To node of link
LENGTH	Length of link
DIRECTION	Direction code: 0, 1 (two-way, one-way A to B)
TYPE	Type of link: 1 – Standard roadway link 2 – Transit only link (bus or rail) 98 – Walk or bicycle only link
FACILITY_TYPE	Type of facility: 1 – Freeway 2 – Major regional arterial 3 – Principal arterial 4 – Minor arterial 5 – Collector 6 – Ramp 9 – TAZ centroid connector
LANES	Number of through lanes
HOV	High-occupancy vehicle code 1 – General purpose lane 2 – 2+ HOV lane 3 – 3+ HOV lane
SPDLMT	Speed limit
FFSPD	Free-flow (zero-volume) speed
BNODE_CTL	Intersection control at B-node 0 – No control 1 – Yield sign 2 – Stop sign 3 – Traffic signal
SCREENLINE	Screenline number for summaries
AWDT	Average weekday traffic count data

Transit networks may include separate network data for local buses, express buses, bus rapid transit (BRT), light-rail transit (LRT), heavy-rail transit (HRT), commuter rail (CR), and in some larger regions, high-speed rail (HST). Transit networks are, in some ways, more complicated than roadway networks. They are represented by links and by itineraries. For buses operating in mixed flow, the state-of-the-practice uses the roadway network to represent mixed flow transit links in order to ensure consistency between maximum speeds used for buses and congested roadway speeds. The roadway network may be supplemented with nonroadway transit links to represent bus only links and lanes as well as links for modes operating on fixed guideways. Transit itineraries list the sequence of transit (roadway) network nodes traversed by each route, whether or not the node represents a transit stop, and headway information for the route.

Nonmotorized mode networks generally represent walk and bicycle modes. These modes may be represented by the roadway network with facilities prohibiting pedestrians and bicycles removed and pedestrian or bicycle only facilities added for the path building and skimming process. The networks and skimming process are commonly based on shortest distances. Nonmotorized travel times for mode choice are commonly estimated using a constant speed for each appropriate nonmotorized mode.

This chapter is organized as follows. Section 3.1 discusses checks of socioeconomic data. Section 3.2 describes checks of transportation network data, while Section 3.3 discusses checks of network skims and path building.

3.1 Socioeconomic Data

3.1.1 Sources of Data

The main sources of data for validation of input socioeconomic data are, primarily, the same sources of data used to develop the data. Few regions have multiple sources of the same socioeconomic data for a particular year. The main sources of socioeconomic data are:

- **Census data –** The decennial U.S. Census provides information on the full set of persons and households in the country and can be summarized at a fine level of geographic resolution, such as the zone level or below. Data from Summary File 3 (SF3) can be used for univariate distributions of household and population data such as households by household size, households by income group, households by structure type, households by auto ownership, and population in households.

- **American Community Survey[6] –** The decennial Census of Population and Housing collects data about the number of people residing in the United States and their relationship within a household, age, race, Hispanic origin (ethnicity), and sex. It also collects information about the number, occupancy status, and tenure (ownership

[6] http://www.census.gov/acs/www/Downloads/ACSPUMS.pdf.

status) of the nation's housing units. In the censuses of 1980, 1990, and 2000, information about topics such as income, education, employment status, disability status, housing value, housing costs, and number of bedrooms were asked on the "long form." Since there is no long form associated with the censuses starting in 2010, data on these topics will come from the American Community Survey (ACS). The ACS is conducted continuously by the Census Bureau, not only at the time of the decennial census, and provides the information formerly included in the long form. The following are some of the data useful to travel forecasting that are available in the ACS:

- Demographic Characteristics:
  - Age;
  - Sex; and
  - Relationship to Householder (e.g., spouse).
- Economic Characteristics:
  - Income;
  - Labor Force Status;
  - Industry, Occupation, and Class of Worker;
  - Place of Work and Journey to Work;
  - Work Status Last Year; and
  - Vehicles Available.
- Financial Characteristics:
  - Tenure (Owner/Renter);
  - Housing Value;
  - Rent; and
  - Selected Monthly Owner Costs.

The decennial census collected detailed population and household data from about one in every six households every 10 years using the long census form. This process is being replaced by the ACS, which samples about one in every 40 addresses every year, or 250,000 addresses every month. This allows the Census Bureau to produce data every year rather than every decade. For areas with large populations (65,000 or more), survey estimates are based on 12 months of ACS data. For all areas with populations of 20,000 or more, the survey estimates are based on three years of ACS data. The Census Bureau is planning to produce estimates for all areas, down to the census tract and block group levels, based on five years of ACS data. The U.S. Census plans to release more ACS data each year as shown in Table 3.3.

Table 3.3 ACS Data Releases

Data Product	Population Threshold	Planned Year of Release				
		2009	2010	2011	2012	2013
1-year Estimates	65,000+	2008	2009	2010	2011	2012
3-year estimates	20,000+	2006-2008	2007-2009	2008-2010	2009-2011	2010-2012
5-year Estimates	All areas[a]	–	2005-2009	2006-2010	2007-2011	2008-2012

Source: U.S. Census Bureau.

[a] Five-year estimates will be available for areas as small as census tracts and block groups.

With the use of the long form in the decennial census ending, Census Transportation Planning Package (CTPP) data are planned to be a product based on the ACS. In addition to providing information on the place of work of residents and the journey to work noted above, the CTPP also provides cross-classifications of socioeconomic data for households at the zone level. An example cross-classification is the number of households by household size by household income.

While the Census Bureau is planning to produce estimates for all areas down to the census tract and block group levels, the estimates will be based on about a 12.5 percent sample of addresses collected over a five-year period. As such, they may not be appropriate for developing or validating input data on a TAZ level, but should provide good summary data for validation checks more aggregate levels. The ACS data may be especially useful for input data validation checks between decennial censuses.

- **Utility hook-up data –** It is sometimes possible to obtain information on households from local utility companies. Data from these sources must be used with care since there is no guarantee that a housing unit is occupied even if it is hooked up to a utility provider, some multiple-unit housing does not have separate utility hook-ups, and it might not always be possible to distinguish households from businesses.

- **School enrollment data –** School enrollment data from public and private schools might be obtained for travel models that explicitly model school related travel such as home-based school and home-based college/university trips or tours.

- **Local land use data and parcel files –** Regions using land use-based travel models must have access to land use data or parcel files. Many jurisdictions have property tax records on-line. With the increasing reliance on Geographic Information System (GIS)-based data storage, such files are becoming more readily available and easier to process. Typically, they are obtained from cities, counties, and other taxing districts.

- **Quarterly Census of Employment and Wages (QCEW) Data –** Employment data are the most difficult data component to collect. One primary source of employment data is the QCEW collected by state Departments of Labor (DOL) for the U.S. DOL. These data have replaced the U.S. DOL Employment, Wages, and Contributions, ES-202 file. Care must be used processing data from this file since the address information shown in the file may not reflect the true work locations of employees. For example, some franchises may list all of the employees at one single location for this file. It might be necessary to sign a confidentiality agreement in order to obtain the data.

- **Market Research Listings –** Many market research firms offer commercial listings of all (or major) employers and number of employees by county and city. The listings may show business locations by P.O. boxes as well as by street addresses. Commercial listings offer these data on a subscriber basis with a range of access and purchase options.

- **Local Area Population and Employment Data –** Many jurisdictions collect and record some type of population data. But few areas record employment data other than a broad listing of the employers with the highest number of employees locally. Chambers of Commerce often publish lists of member businesses.

- **Aerial Photography and Existing Land Use –** Often aerial or satellite photographs available at several locations on the web can be used to update or validate existing land uses. The resolution of the photography can be good enough to differentiate many residential and nonresidential areas. When compared with the aerial photographs, each land use can be associated with a particular land use type (e.g., residential dwelling units, retail and industrial) for each building. It is crucial to know the date of the imagery (when the pictures were taken) prior to using it for land use updates.

3.1.2 Aggregate Checks

The primary aggregate validation checks for socioeconomic data are the summation of TAZ data to different geographic areas and comparison to observed data. Summation of data such as population and households to political divisions such as cities and counties is particularly important, especially if the base year for the input data is close to the decennial census. However, with the release of the ACS data for regions, it will be possible to compare socioeconomic data to Census estimates for areas with 65,000 or more residents every year, areas with 20,000 or more residents based on three-year summaries and averages, and all areas based on five-year summaries and averages. In addition to being able to check aggregate totals of data such as population and households, the ACS data will provide the means to check information such as median incomes and income distributions, household size distributions, and vehicle availability distributions.

The ACS will also provide a means to check employment data. The check will probably be most accurate at the regional level with decreasing levels of confidence for smaller geographic areas. Resident labor force information regarding industry, occupation, and class of worker, coupled with place of work information, can be compared with employment (from establishments) at the regional or subregional level. That is:

$$EstEmp_{Type} = EmpRes_{Type} + EmpNonRes_{Type}$$

Where:

$EstEmp_{Type}$ = The estimated employment by type for the region or subregion;

$EmpRes_{Type}$ = Residents reporting place of work in the region or subregion by type from the ACS; and

$EmpNonRes_{Type}$ = Nonresidents of the region reporting place of work in the region or subregion by type from the ACS.

The employment by type estimated from the ACS data using the above formula should approximately equal the employment estimates for the same geography from the input socioeconomic data. The application of the above check will be easier for more isolated MPOs since the impact of nonresidents of the region working in the region will be minimized.

It is also important to compare regional employment with regional workers. Employment estimates are developed from employer surveys, QCEW data, and other sources focused on businesses and other establishments. Workers are estimated from census data or other sources focused on the residents of the region. While the estimates of employment and workers do not have to exactly match for a region, there should be some consistency between the estimates.

3.1.3 Disaggregate Checks

Multiple independent sources of disaggregate socioeconomic data are not generally available. This may change as five-year ACS estimates of socioeconomic data become more readily available. In some regions, estimates of socioeconomic data for years between census years are made through incremental annual updates to the most recent census data for the region. In those areas, the five-year ACS estimates of the socioeconomic data can be used as independent estimates of the socioeconomic data on a TAZ-by-TAZ basis. However, since the ACS data will be estimates based on an effective 12.5-percent sample of addresses collected over a five-year period, discrepancies between the ACS data and the local estimate of the socioeconomic data might be related to sampling error associated with the ACS data just as easily as an error in the local estimate of the data.

Spot checks of input dwelling unit and household data might also be made using aerial photograph data. While households are not exactly equivalent to dwelling units due to unoccupied units and multiple households living within what appears to be a single unit, there is a high correlation these two variables. To perform the spot check, a random sample of TAZs could be drawn and the dwelling units in the TAZs counted using recent aerial photographs. TAZs with large numbers of multifamily units would need to be skipped due to the difficulty of counting the numbers of dwelling units from aerial photographs. The match between input household data and the counted dwelling units for the sample of TAZs would provide a level of confidence in the coded input data.

Disaggregate checks of employment data can be performed if independent data are available. For example, if the input employment data are developed from QCEW data, detailed checks of the input data might be made using files purchased from a commercial vendor.

3.1.4 Criteria Guidelines

There are no applicable criteria guidelines for aggregate or disaggregate checks of input socioeconomic data.

3.1.5 Reasonableness and Sensitivity Testing

Several types of reasonableness and sensitivity checks for socioeconomic data can be performed. On an aggregate level, a number of regional rates should be calculated and compared to historical data for the modeled region. Tables 3.4 and 3.5 show some national demographic trends over the past 40 years. Similar demographic statistics and trends should be developed for the modeled region. Newly-developed socioeconomic data should fit reasonably well with the trends for the region.

Table 3.4 Summary of Demographic Trends from the NPTS

	Year					
Statistic	**1969**	**1977**	**1983**	**1990**	**2001**	**2008**
Persons per household	3.16	2.83	2.69	2.56	2.63	
Vehicles per household	1.16	1.59	1.68	1.77	1.78	
Workers per household	1.21	1.23	1.21	1.27	1.33	
Vehicles per worker	0.96	1.29	1.39	1.40	1.34	
Vehicles per licensed driver	0.70	0.94	0.98	1.01	1.00	

Source: 1969, 1977, 1983, 1990, 2001, and 2008 NPTS.

Table 3.5 Percent of Households by Vehicles Available

	Year					
Vehicles Available	1969	1977	1983	1990	2001	2008
No vehicles	20.6%	15.3%	13.5%	9.2%	8.1%	
One vehicle	48.4%	34.6%	33.7%	32.8%	32.4%	
Two vehicles	26.4%	34.4%	33.5%	38.4%	40.4%	
Three or more vehicles	4.6%	15.7%	19.2%	19.5%	19.1%	

Source: 1969, 1977, 1983, 1990, 2001, and 2008 NPTS.

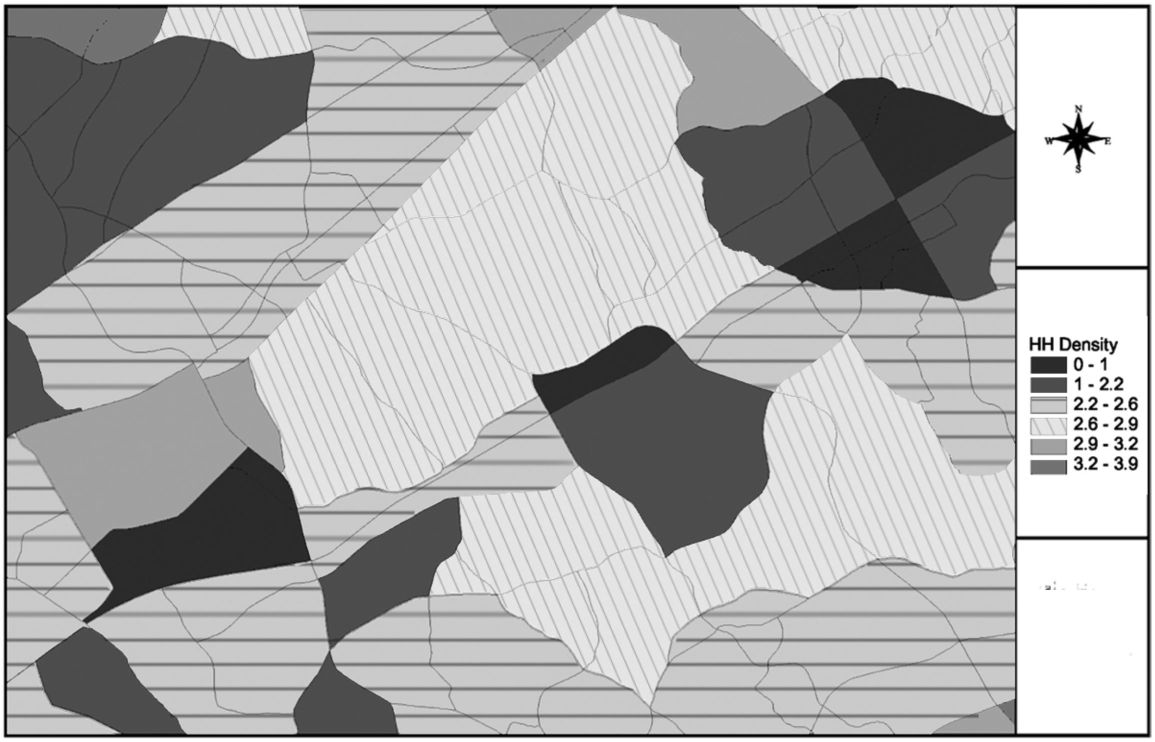
A second type of reasonableness check is the preparation of GIS plots. Almost any district-level or TAZ-level data can be effectively displayed using a GIS. Example zonal socioeconomic data which can be checked using a GIS include population, households, average household size, proportions of households by socioeconomic stratum (e.g., income level or auto ownership), employment, and employment by category.

Two types of checks which can be performed with a GIS include:

- Calculate densities and plot using thematic mapping. Calculate population and employment density in persons per acre (or square mile). Densities should be grouped to produce a reasonable number (e.g., four to six) equal area or equal number of zones categories for the region. Color or shading can be used to convey densities. An example is shown in Figure 3.1.

- Compare existing to most recent year, or forecast year to current year totals by zone or district and plot changes. Plot so that positive and negative changes can be easily identified.

Figure 3.1 Example Socioeconomic Data Thematic Plot for Visual Checking



3.1.6 Troubleshooting Strategies

Table 3.6 shows some of the typical issues that might be found from tests of input socio-economic data.

Table 3.6 Troubleshooting Strategies for Issues with Input Socioeconomic Data

Issue	Potential Troubleshooting Strategies
1. Aggregated socioeconomic data are significantly different from independent data sources	• Check data aggregation procedures to ensure that TAZs were not skipped or double counted • Check for differences between coverage areas for socioeconomic data and independent data • Determine which data set is likely to be more accurate
2. Overall aggregate rates (e.g., average household size) are different from observed data or trends	• Recheck observed data for processing errors • Perform checks for smaller geographic areas to isolate problems or determine if the difference is general in nature
3. Coded input data do not match independent data at a disaggregate level	• Recheck coded input data for processing errors • Check trends in independent data (especially ACS data) over time for consistency
4. Inconsistency (not necessarily inequality) between number of workers and number of jobs at the regional level	• For base year, determine which data source is more reliable and adjust data from the less reliable source to be consistent with it • For forecast year, determine which variable has more reasonable growth rate and adjust growth rate for the variable with the less reasonable rate

3.1.7 Forecasting Checks

Forecasting checks of input socioeconomic data are focused on comparisons to the most recent base year data. The checks should be similar to those described under Reasonableness and Sensitivity Testing (Section 3.1.5).

A basic check is the growth rate in aggregate variables such as population, households, and employment. Typical annual growth rates in population can vary but are usually in the range of zero to two percent and are seldom greater than four percent. Table 3.7 summarizes annual population growth rates for 368 Metropolitan Statistical Areas[7] (MSAs) based on U.S. Census data from 2000 to 2007, classified by region of the country and population range. Not surprisingly, areas in the Northeast and Midwest are growing

[7] The New Orleans MSA is excluded from this analysis.

more slowly than areas in the south and west. On average, larger areas are growing a bit faster than smaller areas. Some other findings from this data set include:

- Fifty-nine of the 368 MSAs lost population from 2000 to 2007. More than one-half of these areas had populations of less than 200,000.
- Eighty-three percent of MSAs had annual growth rates under two percent.
- Only two MSAs had annual growth rates of greater than 5 percent, and they were among the smallest MSAs (the annual rate of 10.7 percent was for the smallest MSA). However, high growth MSAs were found in all population ranges, with three of the top 11 growth rates occurring in MSAs with populations of more than 1,000,000.

Table 3.7 Average Annual Population Growth, 2000-2007, U.S. MSAs

	Population Range				
Region	**> 1,000,000**	**500,000-1,000,000**	**200,000-500,000**	**50,000-200,000**	**All**
Average Growth Rates					
California	1.3%	2.0%	0.8%	1.8%	1.3%
Midwest	0.7%	0.6%	0.6%	0.4%	0.5%
Northeast	0.3%	0.4%	0.7%	0.3%	0.4%
Other West	2.5%	2.1%	2.1%	1.9%	2.0%
South Central	2.1%	1.6%	1.2%	0.5%	1.1%
Southeast	1.5%	2.0%	1.4%	1.3%	1.5%
All	**1.3%**	**1.3%**	**1.1%**	**0.9%**	**1.1%**
Ranges of Growth Rates					
California	0.2%–3.5%	0.8%–2.7%	-0.2%–2.3%	0.9%–2.6%	-0.2%–3.5%
Midwest	-0.3%–1.5%	-0.7%–1.9%	-0.5%–2.9%	-1.0%–1.9%	-1.0%–2.9%
Northeast	-0.5%–1.4%	-0.3%–1.2%	-0.3%–2.4%	-0.8%–2.8%	-0.8%–2.8%
Other West	1.2%–4.5%	0.5%–3.6%	0.3%–4.7%	0.3%–6.7%	0.3%–6.7%
South Central	1.2%–3.8%	0.7%–3.4%	-0.3%–3.5%	-0.8%–1.9%	-0.8%–3.8%
Southeast	0.7%–3.3%	0.8%–4.7%	-0.9%–3.7%	-0.5%–10.7%	-0.9%–10.7%
All	**-0.5%–4.5%**	**-0.7%–4.7%**	**-0.9%–4.7%**	**-1.0%–10.7%**	**-1.0%–10.7%**

It is also critical to check that growth rates for variables, such as population, households, workers, autos, and employment, are consistent with one another. The statistics shown in

Table 3.4 can be computed and compared regionwide for the base and forecast years. Large differences should be checked; if correct, there should be some logical explanation for the difference. The growth rates for workers and employment should be very close, unless there is some mitigating factor (such as a combination of high growth in commercial development inside the modeled region and high growth in residential development outside the region).

Spreadsheets and thematic maps can also be useful in checking growth rates in socioeconomic data for reasonableness. Maps such as the one shown in Figure 3.1 can be developed for growth rates in variables between the base and forecast year.

3.2 Transportation Network Data

3.2.1 Sources of Data

In contrast to socioeconomic data which is relatively volatile over time (at least on a TAZ level), transportation network data remain relatively stable over time. Most models have existing transportation networks that must simply be updated to reflect new roadways, increases in roadway capacity (e.g., through the addition of travel lanes), or new transit services. Thus, the main sources of validation data for checking input transportation network data are not the same sources of data used to develop the data. The main sources of transportation network data are:

- **Topologically Integrated Geographic Encoding and Referencing (TIGER) files –** The TIGER database produced by the Census Bureau contains shapefiles that can be used to create a roadway network. The data file contains visible linear features such as roads, railroads, and hydrography, as well as nonfeature edges, and nonvisible Current boundaries. Additional attribute data associated with the linear features found in the All Lines shapefiles are available in relationship files. The amount of detail available in these databases is more than is necessary for the topology of the model network. Consequently, the user must take care to filter out unwanted detail, such as local streets.

- **Commercial vendors –** With the almost universal use of GIS, easy to use, commercial digitized map files are available from many vendors. Many of these are simply enhanced TIGER files, which save the user the time and effort of editing census TIGER files, but others may provide additional information that can be used to check the roadway network. Examples of data augmenting TIGER files include characteristics such as street width, posted speed, and facility type.

- **Aerial photography –** Often aerial or satellite photographs available at several locations on the web can be used to update or validate roadway networks. The resolution of the photography can be good enough to differentiate network connectivity (e.g., grade separations as opposed to intersections), number of lanes, and locations and types of intersection control (stop signs and signals). Street-level views being offered on some web sites provide an added level of detail in some areas. In these street-level views, channelization elements (pavement markings, islands and signs) are usually visible. Features of traffic control devices, including stop signs and signal heads are often visible.

- **Transit route maps and schedules –** The primary source for transit network data is the route maps and schedules provided by the transit operators. This information may be used for both transit network coding and network validation. Transit schedules and route maps are typically used to develop route itineraries and headways input to the travel models. They may also be used to help develop relationships between bus speeds and roadway speeds for buses operating in mixed flow or transit travel times for transit vehicles operating on exclusive guideways. The data may also be used for validation. For example, modeled end-to-end travel times for bus routes operating in mixed flow may be compared to scheduled times.

- **Transit operation summaries –** Transit operations departments for transit operators typically maintain summaries of their operations including statistics such as number of service vehicle-miles operated by route, number of vehicle-hours operated by route.

3.2.2 Aggregate Checks

As with socioeconomic data, the primary validation checks for input transportation network data are the aggregation of coded network data by various strata for comparison to independently summarized data for the same strata. For example, the coded lane-miles of roadway could be summed by facility type, by speed limit, or by geographic area and compared to similar summaries from available GIS data. For the transit network, coded vehicle-miles and vehicle-hours by route can be estimated from coded routes, headways, and periods of operation and compared to transit operator estimates of the actual service provided. Table 3.8 shows an example of comparisons of model results to transit operator summaries for a modeled region.

Table 3.8 Example Regional Transit Network Coding Check

	Daily Vehicle Trips		Daily Service Miles		Average Route Length	
Service Type	**Regional Summary**	**Model Results**	**Regional Summary**	**Model Results**	**Regional Summary**	**Model Results**
Local	4,229	4,197	57,975	57,500	13.7	13.4
Express	461	470	6,970	7,190	15.1	15.3
Regional	75	78	2,520	2,645	33.6	33.9
LRT	248	248	4,740	4,740	19.1	19.1

3.2.3 Disaggregate Checks

Comprehensive disaggregate checks of transportation networks are not generally feasible since they would, in effect, require the development of a second transportation network from independent data sources. Thus, disaggregate transportation network checks may rely on spot checks of the data. A random sample of coded network links could be selected and certain characteristics verified using aerial photographs.

Visual checks of networks are listed under Section 3.2.5, Reasonableness and Sensitivity Testing, since such checks rarely individually compare all coded network links in the modeled region to the actual network. However, it is feasible to check all links in the region for "exceptional" characteristics. For example, a color coded plot of all coded one-way links in the modeled region with directional arrows shown could be produced. Since there should be a limited number of one-way links in the region, detailed checking could be performed in most areas.

It is also possible to perform checks comparing detailed coding to reasonable ranges. For example, coded link lengths can easily be compared to straight line distances calculated from the coordinates of end nodes of the links. Any links with differences outside of a reasonable tolerance accounting for curves could be flagged and checked for reasonableness.

3.2.4 Criteria Guidelines

There are no applicable criteria guidelines for aggregate or disaggregate checks of input transportation network data.

3.2.5 Reasonableness and Sensitivity Testing

Most travel demand modeling software has GIS or GIS-like capabilities for displaying transportation network data. The first level of reasonableness testing is to produce color-coded plots of network characteristics to check for continuity and reasonableness. Examples of network characteristics to plot include facility type, number of lanes, and speed limits. An example is shown in Figure 3.2.

As discussed in Section 3.2.3, all links coded as one-way should be checked visually. A map using color coding and/or arrows to indicate directions of one-way roadways should be produced. All links coded as one-way should be verified as one-way roadways, and the direction of each link should be verified.

Transit networks should also be plotted and checked. In addition to plotting routes, it might be possible to aggregate route information to links in order to plot information such as transit speeds and number of transit vehicles per hour.

Figure 3.2 Example Highway Network Plot for Visual Checking

Note: This figure is not critical in understanding the concept of visually checking highway networks.

It is also worthwhile to build and plot selected paths through the transportation network. For the roadway network, both shortest free-flow time paths and shortest distance paths can be built.[8] In addition to checking the paths for reasonableness, the paths can be checked against web-based programs that build shortest paths based on their representations of the transportation network. Likewise, some transit operators have web-based applications to build the best transit routes for specified interchanges. Thus, the paths built using the coded transit network can be compared to the transit paths suggested by the transit operator.

3.2.6 Troubleshooting Strategies

Table 3.9 shows some of the typical issues that might be found from tests of input transportation network data.

[8] Shortest distance paths can also be built using the valid nonmotorized mode network.

Table 3.9 Troubleshooting Strategies for Issues with Input Transportation Network Data

Issue	Potential Troubleshooting Strategies
1. Aggregated transportation network data are significantly different from independent data sources	• Check data aggregation procedures to ensure that coded network links were not skipped or double counted • Check for differences between coverage areas for coded transportation network data and independent data • Determine which data set is likely to be more accurate
2. Selected test paths through the network are illogical	• Recheck connector links • Check travel time/speed/distance variables for links along the illogical path and for competing paths • Recheck connectivity of network for unconnected roadway network nodes at same location (i.e., creating a grade separation rather than an intersection); for transit, check to ensure transfers are permissible between routes
3. Coded input data do not match aerial photographs at a disaggregate level	• Recheck coded input data for processing errors • Check dates for aerial photographs for consistency with coded network • Double check for through lanes versus parking lanes

3.2.7 Forecasting Checks

Forecasting checks of input network data will need to be against most recent base year data. The checks should be similar to those described under Reasonableness and Sensitivity Testing (Section 3.2.5). Thematic maps (such as Figure 3.2) can be particularly useful when showing only the changes between the base and forecast years. Any differences that appear on a map can be checked to ensure that they truly represent changes in the roadway network, such as highway upgrades or widenings, new roads, etc. Lists of planned projects from sources such as the region's Transportation Improvement Program or Regional Transportation Plan can be used to ensure that any appropriate highway or transit network changes appear in the forecast year networks.

3.3 Network Skims and Path Building

"Impedance" is used as a variable in model components such as trip distribution (discussed in Chapter 6) and mode choice (discussed in Chapter 7). Impedance is a measure of the "cost" to travel between the origin and destination. It is usually a combination of time, monetary cost, and distance related variables (generally referred to as "level of service" variables). Impedance can range from a very simple measure, such as auto in-vehicle travel time, to a "composite impedance" measure, which can be a combination of time, cost, and distance variables by a variety of modes. The choice of an appropriate impedance measure is critical in obtaining accurate models. A simple measure may be sufficient for modeling trip distribution in a smaller modeled region where nearly all trips are made by automobile, but a more complex variable may be needed in mode choice models for regions with substantial use of alternative transportation modes.

Some impedance measures combine different level of service variables for the same mode (e.g., auto). This is usually done through a linear combination of the component level of service variables, in the same way that utility functions are developed for mode choice models. For example:

$$\text{Impedance} = \text{In–Vehicle Time} + B_1 \times \text{Out–of–Vehicle Time} + B_2 \times \text{Cost} \qquad (3.1)$$

The parameters B_k are estimated in the mode choice model or asserted based on information from other models. In the equation above, B_1 represents the weight at which travelers value out-of-vehicle time relative to in-vehicle time (typically 2 to 4) and $1/B_2$ represents the monetary value of in-vehicle time.

The generation of impedance measures depends on two primary components:

1. **Network development –** As discussed in Section 3.2, the highway network includes representations of the main roadways in the modeled region, along with characteristics of the highway links such as free-flow speeds or travel times, distances, facility types, number of lanes, and capacities. Transit networks generally represent every bus and rail route in the modeled region along with stop locations. Characteristics of routes such as headways, fares, travel times between stops, and access times to and from stops are represented.

2. **Network skimming –** Skimming is the process of determining the times, costs, and distances between each pair of zones for which service is available (generally all pairs of zones for auto). Skims may be computed for each component of impedance (e.g., auto in-vehicle time, transit in-vehicle time, transit wait time, transit walk access time, etc.) using the appropriate network and network characteristics.

The main decisions regarding skimming concern the assumptions, or settings, used to determine the zone-to-zone paths. This is more prominent in the transit network skims since there are multiple transit paths between zones, and the optimal path may vary under different assumptions. For example, say there are two paths, one with 20 minutes in-vehicle time and 10 minutes out-of-vehicle time, and one with 12 minutes in-vehicle time and 15 minutes out-of-vehicle time. If in-vehicle time and out-of-vehicle time are weighted equally, the second path would be optimal, but if out-of-vehicle time is weighted twice as much as in-vehicle time, the first path would be optimal.

The accuracy of skim data is critical in obtaining valid mode choice model results. Skim data accuracy is primarily dependent on the accuracy of the network data themselves (see Section 3.2) and on the assumptions made or settings used in producing the skims. At the same time, the consistency of path building procedures and mode choice is crucial for producing logical results from mode choice models when transportation networks are modified in alternatives testing.

On the highway side, the main assumptions are the definition of "cost" in determining the lowest cost paths. The main component of generalized highway cost is travel time, but other variables, including toll cost and highway distance may be used.

The generalized cost for transit skimming should be consistent with the way in which the transportation level of service variables are used in the mode choice utility function, which is a linear combination of these variables. For example, if the out-of-vehicle time coefficient is twice the in-vehicle time coefficient in the mode choice model, out-of-vehicle time should be weighted at twice the weight of in-vehicle time when performing transit skimming.

Transit skim procedures often include a set of rules defining what constitutes a valid transit path. The rules are often based on available observed travel behavior data, such as transit on-board survey data. Some examples include:

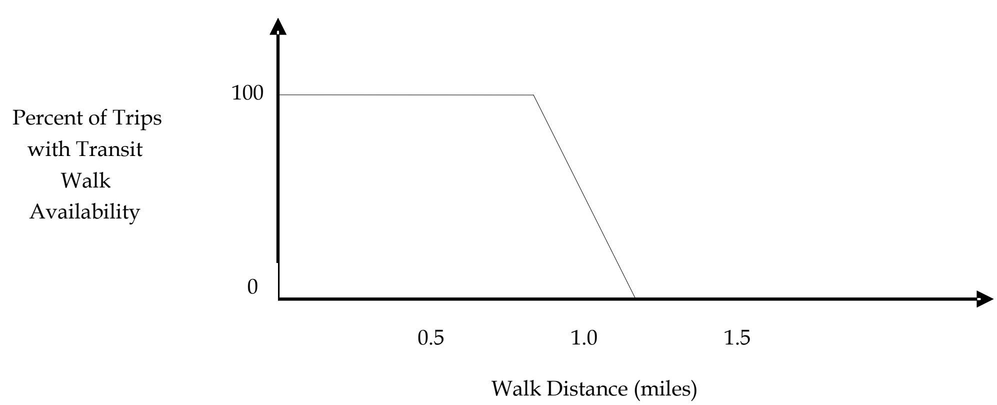
- Maximum walk access distance for transit with walk access modes;
- Maximum walk egress distance for transit with walk egress modes (generally the same limits as for walk access);
- Maximum auto access distance for transit with auto access modes;
- Minimum transit in-vehicle time, to prevent unlikely transit trips of very short distances;
- Maximum number of transit transfers; and
- Maximum total travel time or cost.

It is important to understand the consequences of these types of limits. While some of the limits are set to eliminate only truly invalid paths (for example, an urban model with a rule that total travel time must be less than five hours), others may exclude some borderline reasonable paths. This means that a transit path may be valid in one scenario but invalid in a slightly different scenario to which the results are being compared. For example, if a maximum walk distance of 1.00 mile is used, and a bus stop is 0.98 miles from an activity center in one scenario, but is replaced by a rail station 1.02 miles from the same activity center in another scenario, the latter scenario will not result in any transit riders from the activity center, even if the rail service is far superior to the bus service.

This type of problem is known as a "cliff" because when the transit availability is graphed as a function of the variable being limited, the graph shows a vertical line, or "cliff." The upper portion of Figure 3.3 shows such a cliff graphically. One approach to reducing the impact of the cliff problem is replacement of the hard limit with a piecewise linear function of the availability of transit with walk access as a function of walk distance, as shown in the lower portion of Figure 3.3.

Figure 3.3 "Cliff" in Transit Walk Access Availability

Hard limit: Transit walk access available only if walk distance is less than 1.0 miles



Piecewise linear function : Transit walk access becomes less likely as walk distan ce approaches 1.0 miles

3.3.1 Sources of Data

The network data used for skims are discussed in Section 3.2. A household survey data set, if available, is valuable for validation of highway network skims. In regions with significant transit use, a transit on-board survey data set is a vital source of information for transit skim validation, as well as for mode choice, time-of-day, and transit assignment models. Using a recent on-board survey consistent with base year travel patterns is strongly encouraged in these cases.

3.3.2 Aggregate Checks

Checks of the networks themselves are discussed in Section 3.2. This section discusses checks of skims and path building.

Highway Skims

Whether time, distance, or both are used in impedance measures, it is recommended that skims of both distance and time be created. If toll roads are included in the highway network, cost skims should also be created. Note that some models may explicitly separate drive alone and carpool modes (perhaps multiple carpool modes representing different auto occupancy levels), and there will be skims for each variable for each auto mode. Some models also include separate toll and nontoll alternatives, requiring even more highway skims.

The first tests check the reasonableness of the skims as a whole. Frequency distributions of skims for each variable can be created over all zone interchanges.[9] The key items to review in this distribution are any extremely high or low times, distances, or costs. The extreme values should be similar to those found in the household survey data set. Some interchanges may have observed travel time data to compare with skim values.

Several other reasonableness checks can be performed to ensure that the highway skims include realistic values. The implied speeds for each zone-to-zone interchange can be estimated by dividing the skimmed highway distance by the skimmed highway travel time and converting for units:

$$S_{ij} = \left(\frac{D_{ij}}{T_{ij}}\right) \times 60 \qquad (3.2)$$

---

[9] If the modeling software cannot create this distribution directly, it may be necessary to create a matrix of "1s" and perform a trip length frequency distribution for this "dummy" trip table.

Where:

S_{ij} = Implied speed from zone i to zone j in miles per hour;

D_{ij} = Skimmed distance from zone i to zone j in miles;

T_{ij} = Skimmed time from zone i to zone j in minutes; and

60 = Conversion of minutes to hours.

Once the above calculations are made, several items can be checked, including the minimum and maximum speed by interchange or from a group of zones (e.g., area type), and a simple frequency distribution of speeds on all interchanges. Again, the key items to review in this distribution are the extremes – any very slow or very fast interchange speeds. Some interchanges may have observed average speed data to compare with skim values.

The highway distances can be compared to the straight line distances between zone centroids, computed from the centroids' X-Y coordinates. Generally, the network distances should be somewhat longer, but not much more than 1.5 times as long. Any larger differences may indicate network problems such as connectivity, and paths between such zone interchanges should be manually checked.[10]

These checks should be done for each scenario to which the model is applied. Comparing skims to another "base" scenario – for example comparing forecast year skims to the base year or a "build" scenario to a "no build" – should also be performed. Differences should be directly attributable to the differences in the scenarios' assumptions (e.g., network changes from base to forecast year, specific highway projects in one scenario and not the other).

Transit skims

The transit network is skimmed for all variables used in the impedance measure. It is also necessary to check transit skims used in mode choice that may not be used in trip distribution, and it is helpful to check all skims used throughout the model at the same time. Depending on the number of variables in the model and the number of modal alternatives in the mode choice model, many transit skims may be needed. Transit modes, where available, might include:

- Local bus with walk access;
- Local bus with auto access (park-and-ride, kiss-and-ride);

[10]It is important to account for the fact that the distances associated with the centroid connectors might not be the same as the implied average network distances to reach the highway network from points in the zone.

- Premium bus with walk access;
- Premium bus with auto access;
- Rail with walk access;[11]
- Rail with auto access;
- Bus and rail (both modes used in path) with walk access; and
- Bus and rail (both modes used in path) with auto access.

Variables skimmed for transit networks may include:

- In-vehicle time;
- Transit fare;
- Parking cost (for park-and-ride trips);
- Number of transfers;
- Auto access time (and in some cases, auto egress time);
- Walk access/egress time;
- Wait time;
- Transfer walk time; and
- Transfer wait time.

The last four items comprise "out-of-vehicle time." It is rare for a mode choice model to include all of the variables listed above, but some measures of in-vehicle time, out-of-vehicle time, and cost are typically present.

As with highway skims, the first tests of transit skims are to check their reasonableness. Frequency distributions of the skims for each mode and variable can be created over all zone interchanges. Once again, the key items to review in this distribution are the extremes – any very high or low times, numbers of transfers, or costs. Two additional checks for in-vehicle time are to check skims against route timetables (for present or "backcast" year scenarios) and to compare bus speeds and times to auto speeds and times (from the highway skims). Generally, bus speeds should be somewhat slower than auto speeds due to the need for stops, but speeds might be comparable for express buses. Bus prioritization schemes or exclusive bus lanes may result in higher bus speeds than auto.

As with the highway skims, these checks should be done for each scenario for which the model is applied. Comparison of skims for a scenario to another "base" scenario should be done, and differences should be directly attributable to the differences in the scenarios' assumptions.

[11] Note that there may be several different rail modes in a region, such as commuter rail, subway/elevated, and light rail, that may be modeled separately in mode choice

3.3.3 Disaggregate Checks

There are no applicable disaggregate checks of highway network skim data. However, transit on-board survey can be used for disaggregate validation of transit skims and path building. The paths reported by survey respondents can be compared to paths used in the skimming process. It may not be feasible to individually examine all survey responses, but a sample of survey responses that collectively use all transit routes in the model might be examined. Exact matches between modeled and observed paths are not required; in fact, the survey may reveal many reported transit paths for the same origin-destination pair. The paths between the model and survey may not agree due to:

- Multiple reasonable paths between an origin and destination;
- Differences between individual respondents' sensitivities to components of level of service (e.g., values of time) and the average values assumed by the model;
- Differences between the true starting and ending points of reported trips and the zone centroid locations in the model;
- Unusual paths taken by respondents due to errors they make or unreported circumstances that might change paths (such as changes in travel plans en route); and
- Survey reporting or processing errors.

The modeler should be able to explain any differences based on the above list.

A more automated disaggregate check of transit path-building can be performed if detailed on-board survey data are available. Specifically, the modeled number of boardings for an interchange can be posted on each surveyed trip record. The modeled and reported numbers of boardings can be compared for each survey record and aggregated into a "prediction-success" table. Tables 3.11 and 3.12 show results from a validation effort performed by the Denver Regional Council of Governments. Table 3.11 summarizes results for a specific subgroup of trip interchanges. Observations on the diagonal of the matrix signify trip interchanges where the modeled number of boardings matches the observed number of boardings. Table 3.12 summarizes the results of the prediction success tables for all surveyed trips (linked trips) tested. As can be seen, the transit network and transit path-builder was successful in reproducing the reported numbers of boardings (transfers) for about 67 percent of the surveyed trips.

Table 3.11 Transit Path-Building Prediction Success Table

PM Period Work Trips Using Walk to Rail		Modeled Boardings			
		No Path	1	2	3+
Reported Boardings	No Path	0	0	0	0
	1	7	3	4	0
	2	1	0	0	0
	3+	0	0	0	0

Source: Kurth, et al., Transit Path-Building: "To Multipath or Not to Multipath," 11th TRB National Transportation Planning Applications Conference, Session 15, May 2007.

Table 3.12 Transit Path-Building Prediction Success Results – Simple Path-Builder

		Percent with Skimmed Boardings:		
	Number of Linked Trips	= Reported Boardings	> Reported Boardings	< Reported Boardings
All Trips	1,278	67%	24%	9%
Walk Access	854	67%	23%	9%
Drive Access	424	67%	25%	7%

Source: Kurth, et al., Transit Path-Building: "To Multipath or Not to Multipath," 11th TRB National Transportation Planning Applications Conference, Session 15, May 2007.

3.3.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of skim data.

3.3.5 Reasonableness and Sensitivity Testing

Reasonableness checks have been presented as part of the aggregate checks presented in Section 3.3.2. There are no applicable sensitivity checks of skim data.

3.3.6 Troubleshooting Strategies

Issues with skim data discovered during the checks described in Section 3.3.2 are usually indicative of issues with either the networks themselves or the path building procedures (skim settings). Table 3.13 shows some of the typical problems encountered with highway skims and potential troubleshooting strategies. Table 3.14 shows the same for transit skims.

Table 3.13 Troubleshooting Strategies for Issues with Highway Network Skims

Issue	Potential Troubleshooting Strategies
1. Very long highway trip lengths for some zone interchanges	• Check highway network for improperly coded link distances, times, or speeds (check that units are correct) • Check paths for reasonableness • If paths are circuitous, check links that "should" be traversed for improper coding or lack of connectivity
2. Highway speeds not reasonable	• Check highway network for improperly coded link distances, times, or speeds
3. Highway distances much longer than straight line distances between zone centroids for some zone interchanges	• See checks for Issue 1 above
4. Paths for some zone interchanges are too circuitous	• Check links that "should" be traversed for improper coding or lack of connectivity

Table 3.14 Troubleshooting Strategies for Issues with Transit Network Skims

Issue	Potential Troubleshooting Strategies
1. Very long transit trip lengths for some zone interchanges	• Check transit network for improperly coded link/route distances, times, or speeds (check that units are correct) • Check paths for reasonableness • If paths are circuitous, check links/routes that "should" be traversed for improper coding or lack of connectivity
2. Bus speeds not reasonable or inconsistent with highway speeds	• Check bus network for improperly coded link/route distances, times, or speeds • Incorporate or revise relationship between bus and highway speeds
3. Paths for some zone interchanges are too circuitous	• Check links/routes that "should" be traversed for improper coding or lack of connectivity
4. Paths from model do not match paths from survey well for some zone interchanges	• Check model paths for reasonableness • Check survey paths for reasonableness • If paths are circuitous, check links/routes that "should" be traversed for improper coding or lack of connectivity

4.0 Socioeconomic Models

4.0 Socioeconomic Models

As discussed in Chapter 3, the main inputs to travel demand models are transportation network data and socioeconomic data. The socioeconomic data generally include employment by type and households classified by variables, such as numbers of persons, numbers of workers, income level, and/or numbers of vehicles. In aggregate models, these data are required to be identified for specific geographic subregions, typically TAZs, but in some models, may be required at a very disaggregate level such as parcels.

In traditional trip-based travel models, socioeconomic models have frequently been incorporated in the trip generation step. Socioeconomic input data have typically included variables, such as numbers of households, population, or average household size, median income, and numbers of vehicles or average vehicles per household for each TAZ. The socioeconomic models have then allocated the input variables to the one-, two-, or three-dimensional cross-classification strata required for application of trip generation models for each TAZ in the modeled region. An example of the socioeconomic stratification is number of households by income group and household size.

Disaggregate models applied to each individual or household require synthetic populations as input data. These are usually generated by programs that are part of the model system (which may also require control totals of households or persons for specific geographic subregions as inputs).

Validation of models that generate socioeconomic data inputs is discussed in Section 4.1. Validation of models that synthetically generate populations is discussed in Section 4.2.

4.1 Socioeconomic Models

Socioeconomic models may or may not consider the level of transportation service, or accessibility. There are two major ways in which accessibility may be considered:

1. The locations on residential and commercial development are affected by accessibility to the transportation system. These effects are not directly modeled by a transportation model alone; an integrated transportation-land use model is required. Validation of these types of integrated models is not covered in this manual.

2. Decisions on vehicle ownership/availability may be affected by the accessibility of the highway system as well as the transit system and the quality of service for nonmotorized (walk and bicycle) modes. These effects may be analyzed within the travel

modeling process through the use of socioeconomic models, such as vehicle availability models.

While the use of accessibility-based socioeconomic models has been increasing, some travel demand models may also make use of socioeconomic models that do not consider transportation level of service. This may be done when the necessary inputs for trip production models, such as households cross-classified by specific variables, are not directly forecast. Examples of such models include models that classify households by the number of persons, workers, or children living in the households.

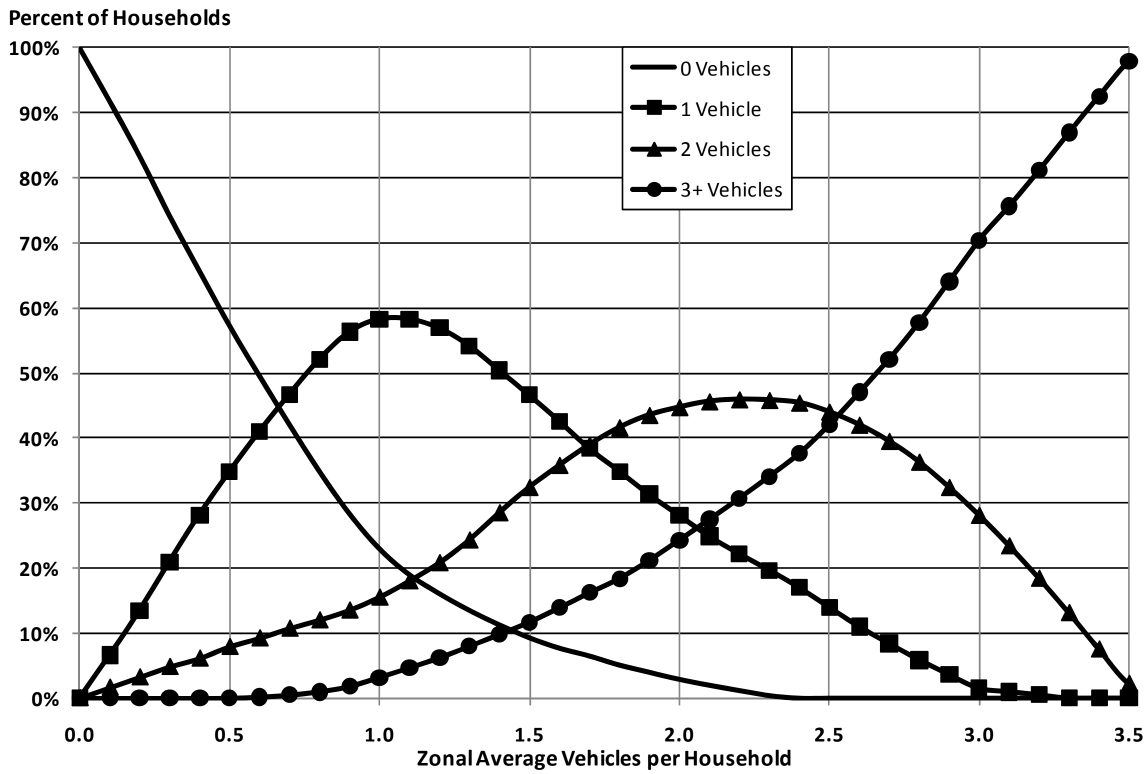
The best state of the practice for socioeconomic models is the use of a discrete choice formulation, usually a multinomial or ordered response logit model, to simulate the "choice" of the number of vehicles (or workers, children, etc.). This type of model can be readily estimated using data from a household activity/travel survey. However, the application of such a model often requires that some of the input variables also be disaggregated. For example, a vehicle availability model may include input variables representing the number of workers and number of persons in the household. If these variables are not forecast in a manner that actually classifies households by level – for example, if the total number of households and total population in each zone is forecast, but not the number of households with one person, two persons, etc. – then some of the inputs for the application of the vehicle availability model are not directly available and may need to be estimated using aggregate share socioeconomic models.

Aggregate share socioeconomic models are used to classify households. The percentages of households in each category are estimated as a function of an average, usually at the zone level. For example, the numbers of households with zero vehicles, one vehicle, two vehicles, etc., are estimated from the average number of vehicles per household in the zone. These models are usually estimated from census data, with curves "smoothed" to produce reasonable results.

Figure 4.1 shows an example of an aggregate share vehicle availability model. This model is applied by reading the shares of each vehicle availability level associated with the average number of vehicles per household, shown along the horizontal axis. For example, if a zone has an average of 1.5 vehicles per household, the model estimates that approximately 9 percent of households own zero vehicles, 47 percent own one vehicle, 32 percent own two vehicles, and 12 percent own three or more vehicles. Note that the curves are estimated so that the sum of the percentages across all vehicle availability levels is 100 percent for each input zonal average vehicles per household.

The discussion of model validation checks for socioeconomic models in this chapter is presented in terms of vehicle availability models. The results can be interpreted similarly for other types of socioeconomic models. For example, the aggregate checks of the numbers of households with zero vehicles, one vehicles, two vehicles, etc., within a subregion could be conducted and interpreted similarly for a model of the numbers of workers (households with zero workers, one worker, two workers, etc.), numbers of households by household size, or numbers of households by income group.

Figure 4.1 Example Aggregate Vehicle Availability Model



4.1.1 Sources of Data

The main sources of data for validation include the following:

- **Household travel/activity survey –** If such a survey is available, it is likely to have been the data source for a choice-based socioeconomic model estimation. It is the best source for information on local household characteristics. It can be expanded to represent the total population and households in a region and can be disaggregated to represent various population segments.

- **Census data –** The decennial U.S. Census provides information on the full set of persons and households in the country and can be summarized at a fine level of geographic resolution, such as the zone level or below. The Census Transportation Planning Package (CTPP), until 2000 based on the "long form" of the census, includes information on a number of cross-classifications of households at the zone level, defined by the MPO or other planning agency. Since the long form was eliminated after the 2000 Census, the Census Bureau has developed the American Community Survey (ACS), which is conducted continuously, not only at the time of the decennial census, and will provide the information formerly provided by the long form.

- **National sources –** Besides the census, relevant national data sources may include the National Household Travel Survey (NHTS) and NCHRP Report 365, Travel Estimation Techniques for Urban Planning, which is being updated (the update is expected to be available in 2010).

4.1.2 Aggregate Checks

The initial checks of aggregate share models are to ensure that the models are internally consistent. For the model shown in Figure 4.1, the sums of the percentages of households by vehicle ownership levels must be 100 percent for each input zonal average vehicles per household. In addition, the internal consistency checks should ensure that the implied output share matches the input share. For the example vehicle share model, the implied average number of vehicles per household resulting from the shares of households by vehicle ownership level for an input zonal average of 1.5 vehicles per household is:

$$\text{Average Vehicles/Household} = (0 \times 0.06) + (1 \times 0.53) + (2 \times 0.32) + (3.67 \times 0.09)$$

$$= 1.5$$

The value used for three or more vehicles in the above check should be reasonable. For example, if the regional average number of vehicles per household for households with three or more vehicles summarized from census data was 3.2, the 3.67 average used above might not be reasonable.

Subsequent to the internal consistency checks, the main aggregate checks of socioeconomic models are comparisons of aggregate model results – for example, the percentage of households by number of vehicles by various market segments – to observed data from the U.S. Census or local household survey. Market segments may be defined by demographic or geographic characteristics, or any other variables by which model results and the comparison data sources are reported. Generally, the census is a good independent data source for validation if the model has been estimated from household survey data.

If a model has been estimated using local household survey data, the model results may be compared to the summaries from the expanded household survey data. Such a comparison can help identify errors in the model estimation and errors in the survey expansion (or differences to be checked between the household characteristics during the survey period compared to the model base year). However, any problems with the survey data set itself, other than in the expansion process, might not be identified since they would exist in both the survey data and the models estimated from the data. It is therefore a good idea to check the model results against census data as well.

Table 4.1 shows an example of an aggregate check of a vehicle availability model. The percentage of vehicles by category and the average number of vehicles per household by county are compared to the observed census data. Some potential model issues can be noted from this comparison. Overall, vehicle ownership is overestimated by about 10 percent, in large part because zero-vehicle households are underestimated by about one third. The overestimation of vehicle ownership is most pronounced in Zeppo County, where the underestimation is about 20 percent. These findings would help the modeler identify how the model might be revised to improve the comparison. (Troubleshooting strategies are discussed in Section 4.1.6.)

Table 4.1 Vehicle Availability Model Results Compared to Observations at County Level

	County							
Variable	**Moe**	**Larry**	**Curly**	**Groucho**	**Chico**	**Harpo**	**Zeppo**	**Region**
Percent 0 Vehicles								
Observed	5%	13%	12%	7%	12%	7%	37%	17%
Model	3%	8%	8%	3%	5%	4%	24%	11%
Percent 1 Vehicle								
Observed	28%	35%	35%	31%	34%	33%	42%	35%
Model	27%	38%	38%	30%	34%	33%	49%	37%
Percent 2 Vehicles								
Observed	44%	37%	38%	44%	38%	43%	18%	34%
Model	47%	39%	39%	45%	43%	43%	21%	36%
Percent 3 Vehicles								
Observed	22%	14%	15%	18%	16%	17%	3%	13%
Model	22%	15%	15%	21%	18%	20%	5%	15%
Average Number of Vehicles								
Observed	1.86	1.55	1.57	1.78	1.58	1.74	0.88	1.44
Model	1.95	1.65	1.65	1.90	1.79	1.82	1.11	1.59

Note that the type of check shown in Table 4.1 can be performed regardless of whether the socioeconomic model is an aggregate share model or a disaggregately estimated logit model.

Another type of validation check, appropriate for validation of an aggregate share model, is shown in Table 4.2. This table shows a comparison between the observed number of households in each vehicle availability category and the number obtained by applying the model using the average vehicles per household for each zone. For example, there are 4,757 households in zones with an average number of vehicles per household of 1.95 to 2.05. From the census data, the numbers of households in these zones owning zero, one, two, and three or more vehicles, respectively, are 605, 1,554, 1,013, and 1,585. Applying the aggregate share model and summing the results over the region yields respective estimates of 590, 1,332, 1,146, and 1,689.

Table 4.2 Check of Aggregate Share Vehicle Availability Model Results

	0 Vehicles		1 Vehicle		2 Vehicles		3+ Vehicles		
Range	**Obs.**	**Model**	**Obs.**	**Model**	**Obs.**	**Model**	**Obs.**	**Model**	**Total**
0.00–0.05	181	181	0	0	0	0	0	0	**181**
0.05–0.15	0	0	0	0	0	0	0	0	**0**
0.15–0.25	0	0	0	0	0	0	0	0	**0**
0.25–0.35	308	293	28	33	5	15	8	8	**349**
0.35–0.45	29	35	17	7	0	3	0	2	**46**
0.45–0.55	382	391	115	119	38	34	33	23	**568**
0.55–0.65	1,531	1,452	515	622	183	190	144	109	**2,373**
0.65–0.75	987	912	425	512	114	150	136	88	**1,662**
0.75–0.85	1,656	1,749	1,211	1,232	472	357	231	232	**3,570**
0.85–0.95	1,208	1,187	938	991	313	320	250	211	**2,709**
0.95–1.05	1,126	1,259	1,263	1,207	555	452	283	310	**3,227**
1.05–1.15	1,178	1,185	1,211	1,280	600	542	396	379	**3,385**
1.15–1.25	2,714	2,569	2,981	3,074	1,279	1,450	1,312	1,193	**8,286**
1.25–1.35	1,373	1,346	1,562	1,710	1,028	886	775	796	**4,738**
1.35–1.45	1,962	1,908	2,570	2,648	1,534	1,481	1,415	1,444	**7,481**
1.45–1.55	2,948	2,886	4,465	4,405	2,367	2,595	2,877	2,772	**12,657**
1.55–1.65	2,431	2,469	4,172	4,124	2,486	2,567	3,076	3,005	**12,165**
1.65–1.75	2,362	2,316	4,251	4,221	2,664	2,805	3,591	3,526	**12,868**
1.75–1.85	1,506	1,425	2,871	2,832	2,030	2,040	2,500	2,610	**8,907**
1.85–1.95	659	708	1,722	1,518	1,102	1,189	1,577	1,645	**5,060**
1.95–2.05	605	590	1,554	1,332	1,013	1,146	1,585	1,689	**4,757**
2.05–2.15	242	201	622	478	289	460	708	722	**1,861**
2.15–2.25	188	172	492	420	439	449	693	770	**1,812**
2.25–2.35	58	19	22	52	45	62	121	113	**246**
2.35–2.45	0	0	0	0	0	0	0	0	**0**
2.45–2.55	17	14	83	42	64	60	93	141	**257**
2.55–2.65	0	9	56	25	42	39	79	104	**177**
2.65–2.75	6	2	12	6	8	10	27	34	**53**
2.75–2.85	0	0	0	0	0	0	0	0	**0**
2.85–2.95	28	6	47	22	49	30	94	160	**218**
2.95 or more	143	10	178	31	87	73	113	406	**521**
Total	**25,828**	**25,296**	**33,383**	**32,940**	**18,806**	**19,408**	**22,117**	**22,490**	**100,134**

Probably the most significant discrepancies between the model results and the observed data in this example occur near the bottom of the table. In the last two rows, for zones with average vehicle availability of 2.85 or higher, the model appears to be significantly overestimating the number of households owning three or more vehicles and underestimating the number of zero and one vehicle households. However, if one computes the observed and modeled vehicles per household for these two categories (assuming 3.5 vehicles per household for the 3+ category), the results are:

- **2.85-2.95:** Modeled – 2.94, Observed – 2.17; and
- **2.95 or more:** Modeled – 3.07, Observed – 1.43.

It is obvious that there are errors in the observed data since the average number of vehicles per household is not close to being within the range for each row. In this case, the observed data needs to be checked and corrected, and the comparison redone.

4.1.3 Disaggregate Checks

For disaggregately estimated socioeconomic models such as logit models, disaggregate validation should be performed in addition to the aggregate checks described in Section 4.1.2. The logit models are estimated with each record in the estimation data set representing a household in the survey data set. There are no applicable disaggregate checks for aggregate share models.

Ideally, disaggregate validation of a model should be performed using a data set that is independent of the data set used for model estimation. Usually, household activity/travel surveys have such small sample sizes that the entire data set is needed for model estimation. Another disaggregate data set available for validation is the U.S. Census Public Use Microdata Sample (PUMS), which also has one record per household. However, the PUMS data do not include geographic resolution below the Public Use Microdata Area (PUMA). PUMAs contain about 100,000 persons and are therefore insufficient for validation of such variables as accessibility, population density, or area type, which are specified at a fairly disaggregate geographic level, such as the zone level.

Limited disaggregate validation can be performed using the same data set used for model estimation, but reporting the results by market segment. Logit model estimation software has the capability to apply the estimated model to a data set that is in the same format as the estimation data set – and therefore to the estimation data set itself. Naturally, the aggregate results of such an application are likely to yield results very similar to the observed choices, but reporting the results by market segment could reveal some potential biases in the model and corresponding areas for improvement. For example, a logit vehicle availability model could be applied to the data set used for estimation, but the results may be reported by income level. It might be found, for example, that zero vehicle households are underestimated in the model for households with higher income levels.

Table 4.3 presents an example of a disaggregate validation check of a vehicle availability model. The results are reported by income level of the household (low, medium, or high) as reported in the model estimation data set. The "observed vehicles" rows refer to the observed number of vehicles in the estimation data set while "modeled vehicles" refer to the results of the application of the model to the estimation data set. The cells where the model results differ from the observed data by more than one standard deviation are shown in bold italics.

Table 4.3 Disaggregate Vehicle Availability Model Check

	Income Category			All Households
	Low	Medium	High	
0 Vehicles				
Observed Vehicles	340	101	65	507
Standard Deviation[a]	18	14	7	32
Modeled Vehicles	**300**	105	63	**468**
1 Vehicle				
Observed Vehicles	684	592	463	1,739
Standard Deviation[a]	32	31	25	60
Modeled Vehicles	653	621	487	1,761
2 Vehicles				
Observed Vehicles	255	492	1113	1,860
Standard Deviation[a]	27	27	25	63
Modeled Vehicles	**300**	**455**	1120	1,875
3+ Vehicles				
Observed Vehicles	66	159	669	894
Standard Deviation[a]	10	20	40	58
Modeled Vehicles	**93**	163	640	896
Total				
Observed Vehicles	1,346	1,344	2,310	5,000
Modeled Vehicles	**1,346**	**1,344**	**2,310**	**5,000**

[a] Standard deviation of the number of vehicles reported in the estimation data set.

Some interesting observations can be made from the model check shown in Table 4.3. The model is underestimating zero-vehicle households by about eight percent. While this could be addressed by increasing the constant for the zero-vehicle alternative relative to the other constants, further examination of the table indicates that this might not improve the model's predictive capability. The underestimation of zero-vehicle households seems to be concentrated in low income households, and so increasing the constant might result in overestimation of zero-vehicle households among the medium and high income categories (and perhaps not enough improvement for low income households). It might be better to look at variables related to low income households, especially since the model seems to be overestimating multiple vehicle availability among low income households.

This type of check could be repeated using segmentation schemes other than income (for example, by geographic subarea or number of workers) to further assist in identifying how the model could be improved.

4.1.4 Criteria Guidelines

There are no specific criteria guidelines associated with the checks in Sections 4.1.2 and 4.1.3. Generally, these are checks of "closeness" between modeled and observed results, and obviously the desire is to have the modeled results be as close as possible to the observed results. Because of differences in the segmentation definitions and in error levels in the observed data, it is impossible to define how close is "close enough." While criteria guidelines are not specified, the validation of socioeconomic models should not be taken lightly since the results impact subsequent models and error may propagate throughout the modeling process.

4.1.5 Reasonableness and Sensitivity Testing

The reasonableness checks for socioeconomic models are mainly the aggregate checks presented in Section 4.1.2.

Sensitivity testing for socioeconomic models may be performed by varying input data and rerunning the models. For example, the percentages of households by income level that are inputs to a vehicle availability model could be revised, and the resulting shares of households by number of vehicles and the average vehicles per household could be compared to the base scenario. Likewise, some vehicle availability models use accessibility to transit as an input variable. For example, as transit accessibility increases, vehicle availability (ownership) decreases. While this might make sense in representing areas with transit oriented development, it would be interesting to note the regional impact of changing all transit accessibility.

Another possible sensitivity test is to use the model to backcast vehicle availability for a past year. Using a decennial census year will likely provide the necessary observed data for comparison and perhaps the necessary cross-classification of households for input data, but other data for the backcast year might be needed as well. For example, if the model has network-based accessibility variables, network data for the backcast year would be needed.

4.1.6 Troubleshooting Strategies

Table 4.4 shows some of the typical issues that might be found from tests of socioeconomic models.

Table 4.4 Troubleshooting Strategies for Issues with Socioeconomic Model Results

Issue	Potential Troubleshooting Strategies
1. Expanded survey data set for estimation and validation is significantly different from independent data sources (e.g., census)	• Check survey expansion factors for consistency with data used for expansion (e.g., census) • Check for differences in socioeconomic data between survey and base years • Determine which data set is likely to be more accurate as a validation data source
2. Overall aggregate results different from observed data	• Recheck observed data for processing errors • Perform checks for market segments and disaggregate validation as applicable • If checks have been done for market segments and appropriate actions taken, consider adjusting constant terms or aggregate shares, depending on model type
3. Results for market segments different from observed data	• Recheck observed data for processing errors • If there is no input variable related to the segmentation variable, consider adding such a variable to the model • Consider revisions to model parameters related to the market segment (e.g., income coefficient for income segmentation)

4.1.7 Forecasting Checks

The forecast year validation checks for socioeconomic models should concentrate on comparisons of the forecast year model results to the base year model results. The base year observed data are no longer directly considered. Unlike the base year comparisons, however, the objective is not to achieve a close match between the forecast and base year results, but rather to ensure that the differences and trends are reasonable. For example, it may be reasonable to expect that vehicle ownership increases somewhat over time, especially if real income is forecasted to increase or growth is concentrated in areas without good transit service.

The main comparisons are similar to those described previously between base year model results and observed data. The comparisons should include both regional and subregional checks. Examples of checks include:

- Numbers of households/persons;
- Average household size;
- Percents of households by number of persons;
- Percents of households by income level;
- Percents of households by number of workers;
- Percents of households by vehicles available;
- Number of employees by employment type;
- Percents of employees by employment type; and
- Regional population/employment ratio.

Particular attention should be paid to those variables and distributions that are used in subsequent travel modeling steps.

Forecast year checks are valuable in finding model issues that may not be evident from checks of base year results compared to observed data. An obvious example is that some problems with socioeconomic forecasts themselves are more easily discovered through checks of model results for a forecast year scenario, compared to the base year. For example, vehicle availability model results may make evident unreasonable differences between the base and forecast years in the socioeconomic data inputs to this model. Another example is an inappropriate level of sensitivity to a model input variable that does not vary substantially in the base year. For example, if the model is too sensitive to income level, and the real income grows significantly between the base and forecast year, the vehicle availability results for the forecast year would likely be unreasonable.

4.2 Synthetic Population Generation

This section deals with validating the results of population synthesis programs, which are used in disaggregately applied models. These programs are referred to as population synthesizers or synthetic population generators. The output of such a program for the base or forecast year is a synthetic population matching available estimates of the modeled region's population, classified by residential location and by household and person characteristics. The output file contains a record for each individual identifying his/her characteristics and those of the household of which the individual is part. Population synthesis is often done using some form of iterative proportional fitting (IPF), matching marginal totals of population/household characteristics. These marginal totals need not be one-dimensional if observed cross-classifications of variables are available.

Person characteristics often used in population synthesizers can include:

- Age;
- Gender;
- Worker status (e.g., full-time, part-time, nonworker);
- Student status (e.g., university/college, secondary school, elementary school, pre-kindergarten, nonstudent); and
- Driver's license status (yes/no).

Each person in the modeled region is simulated as part of a household with one or more persons (or, in some cases, as part of a group quarters living arrangement). Other household-level variables can be derived from the person characteristics, including:

- Number of persons in household;
- Number of workers in household;
- Number of children in household; and
- Number of older people (e.g., age 65 or over) in household.

Certain household characteristics, such as household income, may be used as control totals for the population synthesis program and be assigned to all household members. Other household characteristics, such as the number of vehicles, may be modeled by a socioeconomic model of the type described in Section 4.1 after application of the population synthesizer.

The main inputs to a population synthesizer include control totals for the marginals for various household/person characteristics and a seed distribution of the person and household characteristics for the base year or similar period.

4.2.1 Sources of Data

There are three major data sources that may be relevant to the validation of population synthesizers:

- **Control totals for classification variables –** The best source for control totals for most person and household variables is the U.S. Census, which provides totals for a wide variety of variables at a very fine level of geographic resolution. Relevant census products include summary files, particularly SF1 and SF3 and the Census Transportation Planning Package (CTPP). The CTPP is now derived from the continuously collected American Community Survey (ACS) rather than the long form of the decennial census.

- **U.S. Census PUMS data –** The PUMS data contain one record per person for each household completing a Census "long form." The PUMS data also include one record for each household with the person data being linked to each household. The PUMS data can be used to provide seed distributions for the population synthesizer albeit at fairly coarse geographic levels of resolution.

- **Household activity/travel survey –** The household survey that is used for the estimation of the various resident person travel model components also provides individual records for persons and households. Household surveys have more precise geographic information but represent a much smaller sample of the population than PUMS.

Census data are often used to provide seed data for the population synthesizer and control totals for the marginal distributions of various person and household characteristics. Thus, validation against Census data can be viewed as validating against the data used for model "calibration." However, such validations are useful especially when they are produced for geographic definitions or for variables other than those used to control the population synthesizer.

Household activity/travel survey data are typically expanded to match regional distributions based on Census data for certain characteristics, such as household size, income group, and vehicle ownership. The expansion factors may also include coarse geographic adjustments. Due to the relatively small sample sizes for most household surveys, the person and household characteristics should probably be considered only at a regional or gross subregional level (e.g., area type). In addition, household surveys are subject to nonresponse bias that may affect distributions of characteristics not considered in the expansion factoring (e.g., age distributions of the population). Nevertheless, the expanded data may provide additional estimates of observed distributions of household and person characteristics for a region.

4.2.2 Aggregate Checks

The main validation data source is likely to be census data for the base year. Detailed summaries and cross-classifications are available at fine levels of geographic resolution, such as the zone, tract, or block group for many variables of interest. If the model's base year is not close to a census year; however, summaries may have to be generated as percentages of the population and households for the census year and applied to the total population and households in the base year. This must be done with caution, as the magnitudes of errors resulting from differences between the census and model base year populations will be unknown. The more disaggregate the checks, the larger these errors will be.

Since census data are available at such fine levels of resolution, the choice of geographic resolution levels for validation can depend on whether meaningful conclusions about the match between the synthetic population and the census data can be drawn, given measures of variation in the population. For example, if the percentage of households with a particular characteristic is 40 percent for a zone, and the standard deviation is 50 percent, it is highly probable that the synthetic population will statistically "match" the observed population in this characteristic regardless of the quality of the population synthesizer. However, examining the results for the same characteristic at a coarser geographic level might produce a more meaningful check.

Aggregate validation checks for population synthesizers can be defined as the comparison of marginal or joint distributions of estimated person or household characteristics at various geographic levels. While comparisons to census data may be made at the zone, tract, or block group levels, coarser levels of geographic resolution might be desired to avoid being overwhelmed with numbers. Examples of typical population and household characteristics that can be checked include:

- Number of households/persons;
- Percentages of households by income level;
- Percentages of households by number of persons;
- Percentages of households by number of workers;
- Percentages of households by number of children;
- Percentages of persons by age group;
- Percentages of persons by worker status; and
- Percentages of person by student status.

Comparisons for cross-classifications of these or other variables, if there are observed data available for comparison, may also be performed.

4.2.3 Disaggregate Checks

There are no disaggregate checks similar to those used for discrete choice models for population synthesizers. While the outputs are indeed disaggregate, there is no disaggregate validation data source. However, pseudo-disaggregate checks may be performed by comparing estimated to observed characteristics for population and households at a fine level of geographic detail, such as zones or census tracts. Bowman and Rousseau used this approach for validating the "accuracy and precision" of the Atlanta region population synthesizer:

> *As used here, the word "accuracy" refers to statistical bias; a variable with a nonzero mean percentage difference between the synthetic population and the census validation value is considered inaccurate. The "percentage difference" is that between synthetic value and census value for a single geographic unit (tract, PUMA, county or super-county). The "mean percentage difference" is the average of this difference across all the geographic units in the region. "Precision" refers to statistical variance; a variable with*

a large variance in the difference between the synthetic population and the census validation value is considered imprecise.[12]

4.2.4 Criteria Guidelines

Population synthesizers are a relatively recent type of process for travel models, and as such there are no applicable criteria guidelines for checks of these models.

4.2.5 Reasonableness and Sensitivity Testing

The reasonableness checks for population synthesizers are mainly the aggregate checks presented in Section 4.2.2.

A sensitivity test relevant to population synthesizers would be a backcast of the synthesized population for a year previous to the model's base year. The most likely backcast year would be a past decennial census year (e.g., 1990 or 2000). Control totals at a fine geographic level of detail and cross-classifications from CTPP would be readily available, and, in many areas, the census data would likely have already been processed for use in a previous version of the travel model.

The backcasting would be done by using the control totals from the backcast year and the same seed distribution that would be used for the synthesis of future populations. The synthetic population can be summarized at various geographic levels, similar to the checks described in Section 4.2.2, and compared to the census data for the backcast year.

4.2.6 Troubleshooting Strategies

Table 4.5 shows some typical issues with population synthesizer results and suggested strategies for dealing with them.

[12] Bowman, J., and G. Rousseau, "Validation of Atlanta, Georgia, Regional Commission Population Synthesizer," **Innovations in Travel Demand Modeling, Summary of a Conference, Volume 2: Papers**, Transportation Research Board, Washington, D.C., 2008, pages 54-62.

Table 4.5 Troubleshooting Strategies for Population Synthesizers

Issue	Potential Troubleshooting Strategies
1. Poor match between synthetic population and census data for some classifications	• Check the seed distribution for accuracy • Check error levels for variables at the given geographic level of detail and consider aggregating the level of resolution • Consider using alternative variables for the seed distribution
2. Backcast results do not match census data for the backcast year	• Consider use of a different seed distribution • Consider use of different variables for the IPF process, particularly noting variables with substantial changes between the backcast and base year populations

4.2.7 Forecasting Checks

The main forecast year checks for population synthesizers are comparisons of the synthetic populations for the forecast and base years. These comparisons can be performed at various levels of geographic aggregation, including regional, subregional, district, and zone. The aggregate checks outlined in Section 4.2.2 can be used for regional and subregional checks while the disaggregate checks outlined in Section 4.2.3 can be used for the district and zone-level checks. The main checks are for consistency between the forecast and base years to ensure that forecasted changes in the population are reasonable.

Many population synthesizers do not derive forecast year synthetic populations directly from synthetic populations for earlier years. That is, the population synthesizer does not start from the earlier synthetic population and "age" the population based on the number of years between the earlier and forecast year and predict household formation and dissipation, births, deaths, and migration. Checking the consistency of the population from the base year to the forecast year is, therefore, critical. One way in which population synthesizers attempt to introduce more consistency is to use an existing synthetic population, for a base year or previously synthesized forecast year, in developing the seed distribution for the new forecast year.

The control totals for various socioeconomic variables that are used as inputs to the population synthesizer are exogenous to the population synthesis process. These are, therefore, socioeconomic data inputs to the process and should be checked according to the process outlined in Chapter 3.

5.0 Amount of Travel/Activity

5.0 Amount of Travel/Activity

The first step in the conventional four-step travel demand forecasting process is the generation of the travel activity for the modeled time period (e.g., daily, peak hour, etc.). Average weekday (referred to hereafter as "daily") travel is normally generated although there are models that have focused on shorter time periods. In four-step models, the generation of daily travel is referred to as trip generation. In activity-based models, model components related to the generation of daily travel include the generation of daily activity patterns for households, tour and subtour generation, and intermediate stop generation. All of these types are referred to in this chapter as "trip generation models."

In four-step models, trip generation is the step where the purposes and amount of travel are calculated. Home-based trips are forecast from the home locations to activities outside the home including work, school, shopping, recreation, social, or other activities. The nonwork related trips are often aggregated into one or more nonwork trip purposes. Four-step models forecast trip productions and attractions, with each trip having one production end and one attraction end. Productions are related to the home end of the trip while attractions are related to the nonhome end. For example, on a daily basis, a single worker may generate two home-based work trip productions at home – a trip from home to work and a trip from work to home. At the work location, the same worker would generate two attractions for the same two trips. Trip productions and attractions focus on the locations generating the travel, not the directionality of travel.

Four-step models must also account for nonhome-based trips which do not start or end at the traveler's home. While the definition of which end of a nonhome-based trip is the production end and which is the attraction end is not critical, by convention the origin is usually designated as the production end. Some regions have subdivided nonhome-based trips into work-based other and other-based other trips. In some cases, the work location has been designated the production end of work-based other trips and the nonwork location the attraction end. Models have used different methods for account for nonhome-based trip making. Since nonhome-based trip-making is performed by persons living in households, a common method for accounting for the trips is to generate them at a household level and then allocate the trips to origins and destinations outside the home.

In conventional four-step models, the most common forms of trip production and attraction models are cross-classification and linear regression,[13] with cross-classification more

[13] Cross-classification is actually a specific form of linear regression where the effects of independent variables (vehicle availability, household size, etc.) are allowed to have a general nonlinear effect. An equivalent linear regression formulation would have appropriately defined dummy variables to represent the effect of each combination value of the independent variables.

common for trip production models and regression more commonly used for trip attraction models. In a cross-classification model, the number of trips is estimated for each combination of the values of two or more independent variables. Typically, these are household characteristics such as the number of persons (household size), number of workers, number of vehicles, and income level. Categories nearly always include aggregations of levels (for example, five or more persons). In a regression model, the number of trips is estimated as a linear combination of variables. For trip attraction models, these are usually zone-level variables representing levels of activity in the zone. They generally include the number of households or persons and the number of employees classified by type (for example, retail, service, etc.).

In activity-based modeling, the focus is the generation of daily activities performed by residents and the travel required to perform those activities. Daily activities that take place outside the home are grouped together to form tours. Each home-based tour begins and ends at home and includes one or more stops along the tour, with common practice designating a particular activity as the primary activity of the tour. Thus, a worker making trips to and from work would be represented by one home-based work tour. Tours may be generated from homes with the tour type being determined by the primary activity performed at the location outside the travelers' homes. Subtours may be generated from nonhome locations, typically work locations. Work-based subtours begin and end at the work location.

Each tour may have intermediate stops, beyond the primary activity location, to perform additional activities. For example, a worker might have to drop off a child at day care on the way to work and pick the child up on the way from work to home. In a four-step model, such a pattern would result in two home-based nonwork trip productions (at the home), two home-based nonwork attractions at the day-care, and two nonhome-based trips, one with an origin at the day care and a destination at the work location and one with an origin at the work location and a destination at the day-care. In an activity-based model, such an activity pattern would result in one tour that includes two intermediate stops.

Activity-based models do not compute trip attractions as used in four-step models. The activity locations that would correspond to the attraction ends of trips are estimated during the destination choice (primary activity and intermediate stop) models, which are described in Chapter 6. This means that any checks of trip attractions described in this chapter cannot be performed until after the destination choice models have been estimated. It is still good practice, however, to perform these checks.

Multinomial logit is the most common form of models of tour and activity generation within activity-based models. These may range from relatively simple models of the exact number of tours or subtours to complex models indicating which combinations of activities are undertaken by a person throughout the day. Variables usually include a variety of characteristics of the traveler and his/her household, area or zone characteristics such as residential or employment density, and accessibility variables computed from "logsums" from subsequently applied destination and mode choice models, as well as a variety of constant terms related to individual alternatives (which may represent complex combinations of activities). It is important to recognize that an activity-based model will have several model components related to trip generation, and each of them must be validated individually, as well as the aggregate results of the activity generation process.

Additional components of travel are typically included in the trip generation step including commercial vehicle trips, internal-external trips, and external-external trips. Most current models use conventional approaches such as cross-classification and linear regression to generate travel for these trip purposes.

The common concept for both four-step and activity-based models is that they both generate measurable amounts of activity or travel. In validating trip generation models, past practice has focused on aggregated statistics related to the numbers of trips generated. With activity-based modeling techniques, the need for disaggregate validation testing has become more pronounced. Both aggregate and disaggregate validation checks may be performed for activity-based models. While it may be more difficult or require more innovation, some disaggregate validation tests may also be performed for four-step models. Evaluation and validation of trip- and tour- generation models are important since the information resulting from this step (i.e., trips or tours) provides the basis for all of the subsequent steps. Errors in this step will propagate through the model chain.

■ 5.1 Residential Person Travel

Residential person travel is defined as travel made by residents of the modeled region, within the modeled region. Not included in this definition are trips made with one or both ends outside the modeled region, truck and commercial vehicle trips, and trips made to and from "special generators" even though they are made by residents of the modeled region. These other types of travel are discussed in subsequent sections of this chapter.

In four-step models, trips made for residential person travel are often referred to as "internal-internal" trips to denote both trip ends being located within the modeled region. The outputs of internal-internal trip generation are productions and attractions by trip purpose at the zone level. In activity-based models, this type of travel is often referred to simply as residential or resident travel. Activity-based models generate outputs at a disaggregate level (i.e., for each modeled person). There are different outputs depending on model structure, but usually produce the following:

- Number of activities by purpose;
- Number of tours by purpose;
- Number of stops on each tour; and
- Number of work-based subtours.

Some models may also estimate joint activity or travel participation among household members, and so additional outputs related to this joint participation may be produced.

5.1.1 Sources of Data

Whenever a recent local household travel/activity survey is available, it is the source for the estimation of resident person travel models, whether in a four-step or activity-based modeling context. In some areas, establishment surveys are performed to assist in the development of trip attraction models.

When recent survey data that could be used for model estimation are not available, model parameters such as trip rates may be transferred from another model or borrowed from other data sources. A common source is the National Household Travel Survey (NHTS), a national survey of personal travel last conducted in 2008. Some other national data sources include the NCHRP Report 365, Travel Estimation Techniques for Urban Planning, which is being updated (the update is expected to be available in 2010), and other documents (e.g., TCRP Report 73, Characteristics of Urban Travel Demand). These reports summarize information from the NHTS and from travel models for various types of urban areas and planning contexts.

Models are estimated, of course, using the best available data, whether it is locally collected, transferred, or from national sources. The truly independent sources of trip generation data needed for model validation generally do not exist for the local area, except in the rare instance when a household survey data set is so large that it can be split into estimation and validation data sets. The validation checks described below therefore rely on the use of the estimation data sources themselves. The national sources can be used as reasonableness checks for various aggregate checks. Household survey data sets collected for two different points in time for a region might be used for sensitivity testing.

5.1.2 Aggregate Checks

Four-Step Models

Trip Productions

The main aggregate checks of trip generation models are comparisons of aggregate model results, usually trips per household by purpose by various other market segments, to observed data from the local household survey. Market segments may be defined by demographic or geographic characteristics, or any other variables by which model results and the comparison data sources are reported.

If a model has been estimated using local household survey data, the model results can be compared to the results from the expanded household survey data. This is particularly useful if the comparisons are made using different stratifications of the data. For example, if the trip generation model is based on household size and income group, comparing the results of an application using the base year socioeconomic data to the expanded survey results by area type could produce important insights regarding the validity of the model. Such a comparison can help identify errors in the model estimation and errors in the survey expansion (or differences to be checked between the household characteristics during the survey period compared to the model base year). However, problems with the survey data set itself, outside the expansion, might not be identified since they would exist in both the survey data and the models estimated from the data.

Table 5.1 shows an example home-based nonwork trip production model for a region. Suppose the model was applied to the base year socioeconomic data and the resulting modeled trips by area type were compared to trips from the expanded household survey used to estimate the model. Further, suppose that the modeled productions to the productions from the expanded survey data were as follows:

Area Type	Ratio Modeled/Expanded Productions
CBD	1.10
Fringe	1.05
Urban	1.03
Suburban	.98
Rural	.95
Overall	.99

Such results would suggest an area type or density bias that might be corrected to produce an improved trip generation model.

Since, quite frequently, the same data set must be used for estimation and validation, it is a good idea to check the model results against other data sources, such as the national sources. Differences between the trip generation model results and the national data sources should be explainable by distinguishing local characteristics. Such checks are discussed in Section 5.1.5.

Table 5.1 Example Home-Based Nonwork Trip Production Model

	Household Size				
Income Group	**1**	**2**	**3**	**4**	**5+**
Low (Less than $15,000)	1.27	2.44	4.98	4.98	4.98
Middle ($15,000–$74,999)	1.38	3.39	4.88	7.53	10.28
High ($75,000 or more)	1.59	2.97	4.88	9.54	10.28

Trip Attractions

The types of checks described above are relevant for trip productions since data sources such as the NHTS and local household activity/travel surveys use households as the sampling unit. There are few sources for checks of trip attractions. One check that can be easily done is to compare the number of home-based work attractions to the total employment by zone. Since home-based work attractions include trips both to and from work, summarized at the workplace, one might expect that an upper bound on the ratio of home-based work attractions to employees would be 2.0.[14] However, there are many reasons for every worker not making two home-based work trips on an average weekday, including the following:

- Some workers are not scheduled to work on every weekday, because their regular schedules include weekends, flex-time, or part-time work.
- A percentage of workers are absent on any given weekday, due to vacations, personal days, sick leave, telecommuting, or work-related travel.
- Any journeys to or from work that include stops on the way are usually considered as combinations of home-based nonwork and nonhome-based trips in most four-step models. (This would not be the case when checking work tour destinations in activity-based models.)

These factors result in the number of home-based work trips per employee being much lower than 2.0. One source[15] suggests that a reasonable range is between 1.20 and 1.55 home-based work attractions per employee.

Balancing Trip Productions and Attractions

In theory, the estimated total trip productions must be equal to the total trip attractions for each trip purpose, since each trip has two ends, a production and an attraction. In model application, however, the estimation of trip productions and attractions will not be exactly equal. This can cause issues with the application of trip distribution models since a doubly constrained model will attempt to match both productions and attractions. While the different trip production and attraction models contribute to the imbalance, much of the difference may be explained by the estimates of the socioeconomic data used in model application.

Before checking the balance between productions and attractions, the effects of nonresidential person travel must be considered. If significantly more people from outside the modeled region work, shop, and perform other activities within the region than residents perform these activities outside, there should be more internal-internal attractions than productions, offset by a corresponding surplus of external trip productions over attractions. This imbalance must be carefully computed since many models use vehicle trips for external travel and person trips for residential travel. (External travel is discussed in

---

[14] Cases of workers making more than one round trip between home and work on the same day do occur but are relatively rare. There are far fewer cases than those of the workers not making home-based work trips described later in the paragraph.

[15] Cambridge Systematics, Inc., *FSUTMS-Cube Framework Phase II – Model Calibration and Validation Standards: Model Validation Guidelines and Standards*, prepared for Florida Department of Transportation Systems Planning Office, 2008.

Section 5.2.) Another area of imbalance that must be considered is special generators (discussed in Section 5.4). Special generator trips are often dominated by or are exclusively trip attractions – for example, airports generate no home-based productions – and so the effects of these "extra attractions" must be considered before balancing productions and attractions regionally.

Once these effects have been considered, the balance between productions and attractions can be checked for each trip purpose. Assuming that the production and attraction models have been developed from the same data source, the ratio of regionwide productions to attractions by purpose should fall in the range of 0.90 to 1.10 prior to balancing. For the base year, the balance between productions and attractions is, in effect, a validation measure. If there is not a close match, the reasons for the lack of match should be investigated.

Activity-Based Models

While the aggregate checks of trip rates described above are directly applicable to the results of four-step models, they can also be performed for activity-based models. For activity-based models, it may be necessary to compile and report results that are not directly generated. Specifically, it should be possible to summarize the tour and intermediate stop information to estimate trips comparable to those used in traditional trip-based models. While this might seem to be a step backward, most regions developing activity-based models have a long history of using trip-based models. If overall regional results in terms of trips per household, trips per person, or the shares of trips by traditional trip purposes are substantially different from those estimated using traditional trip-based models, an investigation of the difference might be warranted.

It is necessary to perform additional checks of activity-based model components. As discussed at the beginning of this section, such models will produce, for each person modeled:

- Number of activities by purpose;
- Number of tours by purpose;
- Number of stops on each tour;
- Number of work-based subtours; and
- Joint activity or travel participation (possibly).

These measures can be summarized by market segment (e.g., area type) in a similar manner to the trip rate summaries discussed earlier and compared to local survey data.

Since trip attractions are not directly estimated in activity-based models, there are no checks of "activity attractions" corresponding to trip attraction checks or checks of balancing of productions and attractions at this stage of the modeling process. However, after destination choice modeling and intermediate stop modeling are performed, some tests analogous to trip attraction model checks may be performed. For example, after the application of all tour destination and intermediate stop location choice models, trip tables analogous to trip-based trip tables should be available for processing. It may be possible to aggregate the various tables by purposes similar to those used for trip-based models to summarize "attractions" by purpose. The resulting summaries could be used to estimate implied "trip attraction rates" for some purposes to check the reasonableness of the resulting travel. For example, it might be possible to summarize home-based work half-tours (i.e., those without an intermediate stop on the journey to or from the work location, depending on the direction of the half-tour) by the work zone. The tours could be divided by the total number of employees to estimate the implied trip attractions per employee. For the base year, the implied trip rate should be reasonably close to the trip attraction rate used for trip-based models.

5.1.3 Disaggregate Checks

The trip generation model forms associated with four-step models, such as cross-classification and linear regression models, are applied aggregately, and so there are no applicable disaggregate checks.

Logit models are disaggregately estimated (one record per person/activity/stop), and therefore disaggregate validation is needed for the logit choice models associated with activity-based models, along with the aggregate checks described above.

Ideally, disaggregate validation of a model should be performed using a data set that is independent of the data set used for model estimation. As discussed above, household activity/travel surveys typically have such small sample sizes that the entire data set is needed for model estimation. In most cases, therefore, there is no independent model estimation data set available for validation.

Limited disaggregate validation can be performed using the same data set used for model estimation, but reporting the results by market segment. Logit model estimation software has the capability to apply the estimated model to a data set in the same form as the estimation data set. For example, a logit model could be applied to the data set used for estimation but the results may be reported by vehicle availability level. It might be found, for example, that certain activity patterns are not being chosen often enough in the model for households with zero vehicles.

5.1.4 Criteria Guidelines

There are no specific criteria guidelines associated with the trip generation checks described above. While no specific guidelines are associated with trip generation, trip generation is the first basic step in the modeling process; errors and inaccuracies in this step may propagate throughout the modeling process.

5.1.5 Reasonableness and Sensitivity Testing

The primary reasonableness checks for trip generation models are comparisons of aggregate trip rates to those estimated for other regions. Table 5.2 shows total person trips per household by trip purpose for different metropolitan area sizes based on the 2001 NHTS. This table is segmented by household size (number of persons in the household). Table 5.2 provides the opportunity for comparisons of several types of trip production model results, including:

- Total trips per household;
- Total trips per household by purpose;
- Percentage of trips by purpose; and
- Trips by purpose for household size categories (1, 2, 3, etc.), if the trip generation model is segmented by household size.

Tables 5.3 through 5.5 show the same type of summaries of the 2001 NHTS data, segmented by number of workers, number of vehicles, and income level, respectively.

It should be noted that the data shown in Tables 5.2 through 5.5 are for all person trips, including both motorized (auto, transit) and nonmotorized (walk, bicycle) trips. It is recognized that many models include only motorized trips. To allow for better trip comparisons for these models, Table 5.6 presents the share of trips that are motorized, by urban area size and trip purpose, from the 2001 NHTS. These percentages can be used to adjust the values in Tables 5.2 through 5.5.

An additional reasonableness check for cross-classification models is to ensure that the rates for individual cells are consistent with one another. This includes checking the following:

- The direction (increase/decrease) between trip rates in adjacent cells along both dimensions is correct. For example, for home-based work trips, the trip rate should be higher for a greater number of workers, holding the other variable constant. However, caution should be exercised since it may not always be correct that a higher value for a variable will result in an increase in the trip rate. As an example, a two person, one worker household might make more nonwork trips than a two person two worker household.

Table 5.2 Trip Rates by Purpose Stratified by Number of Persons by MSA Population

	Number of Persons					All Households
	1	2	3	4	5+	
Home-Based Work						
MSA population greater than 3 million	0.54	1.45	2.16	2.16	2.39	1.54
MSA population between 1 and 3 million	0.56	1.58	2.11	2.17	2.32	1.55
MSA population between 500,000 and 1 million	0.51	1.48	1.90	2.22	2.47	1.52
MSA population between 250,000 and 500,000	0.47	1.44	1.94	2.37	2.34	1.51
MSA population less than 250,000	0.51	1.42	2.08	2.15	2.50	1.48
Not in MSA	0.39	1.35	2.00	2.17	2.51	1.43
Home-Based Nonwork						
MSA population greater than 3 million	1.58	4.20	6.43	9.84	14.01	5.84
MSA population between 1 and 3 million	1.66	4.21	6.64	10.15	13.19	5.60
MSA population between 500,000 and 1 million	1.65	4.58	6.46	9.84	13.26	5.85
MSA population between 250,000 and 500,000	1.71	4.39	6.47	9.49	14.46	5.80
MSA population less than 250,000	1.84	4.38	6.52	10.82	12.86	5.63
Not in MSA	1.59	4.20	6.39	9.30	12.23	5.33
Nonhome-Based						
MSA population greater than 3 million	1.24	2.98	3.84	4.83	5.81	3.27
MSA population between 1 and 3 million	1.19	3.09	4.40	5.30	6.57	3.44
MSA population between 500,000 and 1 million	1.18	2.97	4.05	5.37	7.04	3.48
MSA population between 250,000 and 500,000	1.24	2.99	4.39	5.16	7.42	3.55
MSA population less than 250,000	1.37	3.09	4.39	6.46	6.80	3.61
Not in MSA	1.18	2.93	4.44	5.52	6.69	3.43
All Trip Purposes						
MSA population greater than 3 million	3.36	8.63	12.43	16.83	22.21	10.65
MSA population between 1 and 3 million	3.41	8.88	13.15	17.52	22.08	10.59
MSA population between 500,000 and 1 million	3.34	9.03	12.41	17.43	22.77	10.85
MSA population between 250,000 and 500,000	3.42	8.82	12.80	17.02	24.22	10.84
MSA population less than 250,000	3.76	8.89	12.99	19.43	22.16	10.70
Not in MSA	3.16	8.48	12.83	16.99	21.43	10.17

Source: 2001 NHTS.

Table 5.3 Trip Rates by Purpose Stratified by Number of Workers by MSA Population

	Number of Workers			
	0	1	2	3+
Home-Based Work				
MSA population greater than 3 million	0.02	1.10	2.30	3.94
MSA population between 1 and 3 million	0.02	1.05	2.39	3.85
MSA population between 500,000 and 1 million	0.01	1.12	2.30	3.76
MSA population between 250,000 and 500,000	0.01	1.10	2.40	3.90
MSA population less than 250,000	0.02	1.12	2.44	3.57
Not in MSA	0.02	1.09	2.41	3.88
Home-Based Nonwork				
MSA population greater than 3 million	3.57	5.38	6.92	8.77
MSA population between 1 and 3 million	3.62	4.87	6.65	8.97
MSA population between 500,000 and 1 million	3.62	5.08	6.97	9.89
MSA population between 250,000 and 500,000	3.94	5.17	6.48	11.15
MSA population less than 250,000	3.86	4.97	6.91	8.79
Not in MSA	3.52	4.90	6.57	8.66
Nonhome-Based				
MSA population greater than 3 million	1.46	2.84	4.34	5.04
MSA population between 1 and 3 million	1.62	3.00	4.40	5.61
MSA population between 500,000 and 1 million	1.38	3.01	4.68	5.87
MSA population between 250,000 and 500,000	1.61	3.09	4.53	7.05
MSA population less than 250,000	1.74	3.22	4.82	6.14
Not in MSA	1.64	3.11	4.59	6.50
All Trip Purposes				
MSA population greater than 3 million				
MSA population between 1 and 3 million	5.05	9.32	13.56	17.75
MSA population between 500,000 and 1 million	5.26	8.92	13.44	18.43
MSA population between 250,000 and 500,000	5.01	9.21	13.95	19.52
MSA population less than 250,000	5.56	9.36	13.41	22.10
Not in MSA	5.62	9.31	14.17	18.50

Source: 2001 NHTS.

Table 5.4 Trip Rates by Purpose Stratified by Number of Autos by MSA Population

	Number of Autos			
	0	1	2	3+
Home-Based Work				
MSA population greater than 3 million	0.82	0.95	1.79	2.58
MSA population between 1 and 3 million	0.47	0.83	1.79	2.53
MSA population between 500,000 and 1 million	0.41	0.82	1.74	2.39
MSA population between 250,000 and 500,000	0.35	0.77	1.74	2.35
MSA population less than 250,000	0.58	0.72	1.67	2.33
Not in MSA	0.23	0.63	1.52	2.30
Home-Based Nonwork				
MSA population greater than 3 million	3.15	3.98	7.31	8.09
MSA population between 1 and 3 million	2.47	3.72	6.57	7.57
MSA population between 500,000 and 1 million	2.53	4.07	6.96	7.40
MSA population between 250,000 and 500,000	2.11	3.56	6.68	8.07
MSA population less than 250,000	1.78	4.03	6.18	7.66
Not in MSA	2.29	3.53	5.97	6.84
Nonhome-Based				
MSA population greater than 3 million	1.63	2.19	4.01	4.83
MSA population between 1 and 3 million	1.19	2.19	4.04	4.89
MSA population between 500,000 and 1 million	0.93	2.05	4.30	4.79
MSA population between 250,000 and 500,000	0.94	2.18	3.86	5.35
MSA population less than 250,000	0.85	2.60	3.89	5.06
Not in MSA	0.96	2.11	3.69	4.84
All Trip Purposes				
MSA population greater than 3 million	5.60	7.12	13.11	15.50
MSA population between 1 and 3 million	4.13	6.74	12.40	14.99
MSA population between 500,000 and 1 million	3.87	6.94	13.00	14.58
MSA population between 250,000 and 500,000	3.40	6.51	12.28	15.77
MSA population less than 250,000	3.21	7.35	11.74	15.05
Not in MSA	3.48	6.27	11.18	13.98

Source: 2001 NHTS.

Table 5.5 Trip Rates by Purpose Stratified by Income Level by MSA Population

	Income Level (in 2000 Dollars)					
	Missing	$0-$10,000	$10,000-$25,000	$25,000-$50,000	$50,000-$100,000	Over $100,000
Home-Based Work						
MSA population greater than 3 million	1.11	0.78	1.08	1.43	1.98	2.04
MSA population between 1 and 3 million	0.93	0.60	0.91	1.51	2.24	2.22
MSA population between 500,000 and 1 million	0.89	0.61	1.07	1.59	2.12	2.10
MSA population between 250,000 and 500,000	0.94	0.81	0.94	1.59	2.24	1.96
MSA population less than 250,000	0.78	0.49	0.99	1.67	2.18	2.02
Not in MSA	0.75	0.36	0.91	1.74	2.27	2.02
Home-Based Nonwork						
MSA population greater than 3 million	4.38	4.47	4.63	5.34	7.00	7.38
MSA population between 1 and 3 million	3.97	3.69	3.94	5.24	7.23	7.99
MSA population between 500,000 and 1 million	3.33	4.04	4.90	5.69	7.56	7.76
MSA population between 250,000 and 500,000	3.88	3.97	4.48	6.36	7.34	6.49
MSA population less than 250,000	3.78	4.34	4.54	5.94	6.72	8.55
Not in MSA	3.63	3.31	4.34	5.95	6.95	6.99
Nonhome-Based						
MSA population greater than 3 million	2.00	1.69	2.21	2.96	3.97	5.08
MSA population between 1 and 3 million	2.36	1.45	2.11	3.13	4.80	5.55
MSA population between 500,000 and 1 million	1.81	2.18	2.09	3.18	5.19	5.58
MSA population between 250,000 and 500,000	2.10	2.06	2.46	3.83	4.79	4.85
MSA population less than 250,000	1.90	2.25	2.60	3.67	5.23	5.63
Not in MSA	2.18	1.43	2.68	3.77	4.93	5.90
All Trip Purposes						
MSA population greater than 3 million	7.49	6.94	7.92	9.73	12.95	14.50
MSA population between 1 and 3 million	7.26	5.74	6.96	9.88	14.27	15.76
MSA population between 500,000 and 1 million	6.03	6.83	8.06	10.46	14.87	15.44
MSA population between 250,000 and 500,000	6.92	6.84	7.88	11.78	14.37	13.30
MSA population less than 250,000	6.46	7.08	8.13	11.28	14.13	16.20
Not in MSA	6.56	5.10	7.93	11.46	14.15	14.91

Source: 2001 NHTS.

Table 5.6 Motorized Trip Percentages by Urban Area Population

	Home-Based Work	Home-Based Nonwork	Nonhome-Based	All Trips
MSA population greater than 3 million	96.3%	83.7%	87.8%	86.8%
MSA population between 1 and 3 million	97.4%	87.9%	93.5%	91.1%
MSA population between 500,000 and 1 million	97.6%	89.6%	95.2%	92.5%
MSA population between 250,000 and 500,000	97.8%	89.3%	94.1%	92.1%
MSA population less than 250,000	97.2%	89.5%	94.5%	92.2%
Not in MSA	97.2%	91.1%	94.2%	93.0%
All Areas	**97.0%**	**87.3%**	**92.0%**	**90.2%**

Source: 2001 NHTS.

- The incremental differences between trip rates in adjacent cells make sense. For example, if household size is one of the variables, the increments between one and two person households, two and three person households, etc., should be reasonable in terms of the additional trips adding a household member would produce.

Sensitivity testing for aggregate trip generation models is not considered to be very useful. Cross-classification models already use separate trip production rates for different levels of the input variables. For example, if one of the input variables is household size, the check of the aggregate trip rates by number of persons in the household (the model results that are compared to the numbers in Table 5.2) already show how the model results vary by number of persons. For regression models, the coefficients themselves indicate the sensitivity of trip generation with respect to the input variables.

The logit models used for trip generation in activity-based models are sensitive variables other than socioeconomic characteristics, through the use of logsum terms that represent accessibility. It is therefore possible to conduct tests of the sensitivity of these models to transportation level of service. For example, highway travel times could be increased or decreased by a fixed percentage, and the resulting changes in the number of activities, tours, and stops can be examined.

Generally, speaking, however, trip generation has not been found to be very sensitive to changes in transportation accessibility, beyond the effects of changes in land use and development. Destination, time-of-day, and route choices are much more sensitive to changes in transportation level of service. So a reasonable result might be that changes in trip generation resulting from changes in level of service are small.

5.1.6 Troubleshooting Strategies

In the past, while checks of trip generation results were relatively uncommon, calibration adjustments to trip rates were often made in response to other system-level validation checks. For example, it was sometimes found that vehicle miles traveled (VMT) from the model were too low compared to observed data from traffic counts. A common response would be to increase trip rates for nonwork travel, on the assumption that nonwork trips were more likely to be underreported in the surveys from which the models were estimated. There seems to be some merit to this claim as survey methods have improved over the years, and fewer trips go unreported. However, there was no direct connection between the underreporting of trips and the shortfall in modeled VMT.

Since there is no independent local data source with which to validate trip generation models, the possible model problems that could be indicated by the checks described in this chapter have to do with discrepancies within the survey data set, between the survey data and the socioeconomic data used to apply the model, and between local data and national sources. Table 5.7 shows some of the typical issues that might be found from these tests.

Table 5.7 Troubleshooting Strategies for Issues with Trip Generation Model Results

Issue	Potential Troubleshooting Strategies
1. Total trips from base year model results inconsistent with expanded survey data	• Check survey expansion factors for consistency with model application data • Check for differences in socioeconomic data between survey and base years • Recheck estimated model parameters
2. Trip rates inconsistent across variables in cross-classification model	• Recheck inconsistent rates • Check error levels for estimated rates • "Smooth" trip rates by combining cells in cross-classification
3. Model results inconsistent with national sources	• Recheck estimated model parameters • Check for ways in which local travel characteristics differ than national • Adjust parameters if they seem erroneous
4. Imbalance between modeled productions and attractions by trip purpose	• Check consistency of survey data with model application data • Check to ensure that external and special generator trips have been correctly considered

5.1.7 Forecast Checks

Trip generation forecasts should be compared to base year estimates for reasonableness. Specifically, trip or activity generation rates per household or per person for the forecast year should be compared to the base year. For trip-based models, past trends in many areas, as documented in the 1997 FHWA Model Validation and Reasonableness Checking Manual, have shown total trips per person to be increasing over time, albeit at a decreasing rate. The changes in per capita trip rates are probably caused by a number of factors that may or may not be included in the trip generation models, including:

- Decreasing household sizes since smaller households tend to generate more trips per capita than larger households;
- Increasing mobility through increasing income or auto ownership; and
- Increasing activity participation (i.e., less time spent "at home").

In addition to checking implied per capita trip and activity generation rates, the distributions of trips or activities by purpose should be compared to base year conditions. Again, typical trends in the U.S. have shown an increase in the proportion of nonhome-based travel with commensurate decreases in home-based travel.

For trip-based models, production-attraction balancing factors should also be checked for reasonableness. Large changes from base year factors might suggest an imbalance in socioeconomic forecasts. For example, if home-based work trip attractions had to be substantially scaled downward to match home-based work productions, the forecast employment or workers per household should be reviewed for reasonableness.

Large changes in per capita trip rates or large changes in the proportions of trips by purpose should be carefully reviewed. All changes in per capita trip rates and proportions of trips per purpose should have plausible explanations.

■ 5.2 External/Nonresident Travel

This section deals with personal travel in the modeled region not made by the region's residents or which enters or leaves the modeled region. Components of this segment of travel include:

- **Trips made by the modeled region's residents which leave/return to the modeled region –** These trips may be modeled, based on household survey data, as part of the resident trip modeling process, either with a separate "internal-external" trip purpose (or purposes) or combined with other resident internal-internal trips, in which case the external zones are valid choices during trip distribution. Alternately, these trips may be combined with trips made by nonresidents that enter the region and modeled as a separate trip purpose or purposes, based on external cordon survey data. In this case, it is necessary to remove internal-external trips from the household survey data prior to use for the development and validation of the internal-internal models. This travel segment is hereafter referred to as resident internal-external (I-X).

- **Trips made by nonresidents entering/leaving the modeled region –** Since information on travel by nonresidents is not included in household surveys, data for estimating and validating models must be obtained elsewhere, often from external cordon surveys. A separate trip purpose (or purposes) is defined, representing the specific trips where these nonresidents enter or leave the modeled region. This travel segment is hereafter referred to as nonresident external-internal (X-I).

  In most regions, the percentage of trips entering or leaving the modeled region by a mode other than private auto is miniscule. In many models, therefore, most or all external travel is modeled as vehicle trips, rather than person trips. Trip productions and attractions represent vehicle trip ends, and there is no mode choice component applied for these trips. It is necessary in these cases to consider vehicle occupancy when dealing with person trip components (for example, resident I-X trips from the household survey).

  It should be noted that trips made to and from the airport by visitors who arrive in the modeled region by air are usually considered to be generated at the airport rather than outside the modeled region, and are not considered external trips. These trips are usually handled by airport modeling procedures (see Section 5.4).

  Nonresident X-I productions and attractions must be generated for both ends of the trip, the "internal" end of the trip (produced by or attracted to an internal zone within the modeled region) and the external end (produced by or attracted to an external zone). The total number of vehicle trips (resident I-X, nonresident X-I, and X-X) generated at an external station in the base year is equal to the traffic count at the external station. Generally, this count acts as a control total for the (vehicle) trips generated at the external zone, with the trips categorized by type based on the results of the external cordon survey. When external transit trips are included, the control total (in person trips) is the transit passenger count plus the traffic count multiplied by the vehicle occupancy (from the external cordon survey, or other source).

  Trip productions and attractions for external trips at internal zones may be generated by separate model components that consider the location of the zone relative to the regional boundary. It is important to ensure that such trips are not "double counted" by internal-internal trip generation models.

- **Trips made by nonresidents that have both ends within the modeled region –** This segment consists of travel made by nonresidents excluding the trip to enter or leave the modeled region. These generally include trips made within the modeled region while the traveler is visiting. These trips are excluded from both household surveys (since the travelers are nonresidents) and external cordon surveys (since the trips do not enter or leave the modeled region). This travel segment is hereafter referred to as nonresident internal-internal (I-I).

Many models do not explicitly model nonresident I-I trips, simply including them within the resident nonhome-based trips (even though the survey data on which models of these trips are based does not include nonresident trips). Visitor surveys are not commonly conducted except in areas with large numbers of visitors. In some areas, nonresident I-I trips are modeled as a percentage of nonresident external-internal (X-I) trips. The percentage might be derived from the percentage of nonhome-based trips for residents of the modeled region, based on the household survey data.

- **Trips with both ends outside the modeled region that pass through the region –** These are made by travelers that are not performing any activities within the modeled region but whose travel paths use roadways within the modeled region. This travel segment is hereafter referred to as external-external (X-X).

  External-external trips are usually modeled using fixed trip tables. Base year trip tables are developed by expanding the X-X trips from external cordon surveys while forecast year trip tables are usually factored from the base year tables.

  With the exception of resident I-X trips, activity-based models do not have the data needed to model full activity patterns and tours associated with external trips. It is therefore usual practice to model external travel using a trip-based approach, similar to the approaches used in four-step models. The model validation checks described in this section are therefore relevant to external travel models for both four-step and activity-based models.

5.2.1 Sources of Data

There are four major data sources that may be relevant to the modeling and validation of external/nonresident travel:

1. **Household activity/travel survey –** The household survey used for the estimation and validation of the resident person travel models includes all trips made by residents that leave the modeled region.

2. **External cordon survey –** An external cordon survey is conducted either by intercepting and interviewing travelers as they enter or leave the modeled region or by recording license plate information and mailing a survey form. It includes all travelers that pass through the survey location regardless of resident status or the origin and destination of the trip and is usually conducted for each major external station. Only information about the trip that is intercepted or recorded is collected, but information about traveler and trip characteristics is collected (though these surveys often collect less information than household or visitor surveys). In many cases, some external stations, especially those with low volumes, are not surveyed, and information on origins/destinations, external trip types, traveler characteristics, and vehicle occupancy obtained from survey results at other locations.

3. **Visitor travel survey –** This survey is usually conducted where visitors to a region stay overnight (such as hotels) and asks about all travel made by the respondent (and

perhaps his/her travel party) within a set period, say the previous 24 hours. A complete set of traveler/trip characteristics is obtained, but the sample is limited to people who are staying at hotels. Trips made by visitors who stay somewhere outside the sampling frame or who do not stay overnight are not included.

4. **Traffic/passenger counts at the modeled region's boundary –** Traffic counts at external stations can provide control totals for the number of vehicle trips entering and leaving the modeled region. If a significant number of travelers enter and leave the modeled region via transit, passenger counts at the regional boundary may also be useful in providing control totals for external travel.

Table 5.8 summarizes how these data sources relate to the four categories of external/nonresident travel.

Table 5.8 External/Nonresident Travel Data Sources

	Data Source			
	Household Activity/ Travel Survey	**External Cordon Survey**	**Visitor Travel Survey**	**Traffic/Passenger Counts**
Resident I-X	Included	Included		Volume only[a]
Nonresident X-I		Included		Volume only[a]
Nonresident I-I			Some included	
External-external		Included		Volume only[a]

[a] Traffic/passenger counts include these trips, but do not separate them into their component categories and include no information on traveler or trip characteristics.

5.2.2 Aggregate Checks

Because both traffic (and perhaps transit passenger) count data and external cordon survey data (where available) are used in the estimation of external travel models, there are likely no independent data sources with which to validate these models. The main checks include ensuring consistency with the model estimation data. These checks include the following:

- Comparing total modeled base year vehicle trips (including resident I-X, nonresident X-I, and X-X) trips to the base year traffic count for each external station. Caution should be exercised in considering that traffic counts may include trucks, which are modeled separately (see Section 5.3). The best approach is to compare the vehicle trips associated with external/nonresident passenger travel to the traffic count excluding trucks, and to compare separately the truck model results with the truck count.

- Comparing total implied trip rates, including all internal and external passenger travel, across zones. The implied trip rates are the total person trip productions (internal and external) per household and total person trip attractions per employee. Particular attention should be paid to comparing implied trip rates for zones located near the modeled region boundary and interior zones. Because models should generate more external travel for outlying zones, this check will ensure that these additional external trips were not "double counted" along with resident internal-internal trips. A corresponding check is the percentage of total generated person trips that are external, which should be higher for outlying zones.

5.2.3 Disaggregate Checks

There are no applicable disaggregate checks of external travel models.

5.2.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of external travel models.

5.2.5 Reasonableness and Sensitivity Testing

The reasonableness checks for external travel models are mainly the aggregate checks presented in Section 5.2.2.

5.2.6 Troubleshooting Strategies

Table 5.9 shows some typical issues with external travel model results and suggested strategies for dealing with them.

5.2.7 Forecasting Checks

Forecast internal-external and external-external travel should be compared to base year for reasonableness. Growth rates in travel can be compared to growth rates for internal person trips, and significant differences should be explainable using demographic forecast data.

Table 5.9 Troubleshooting Strategies for Issues with External Travel Model Results

Issue	Potential Troubleshooting Strategies
1. Total vehicle trips for each external station for base year model inconsistent with the corresponding traffic count	• Check traffic count data for errors, seasonal adjustment, and adjustment between the count and model years • Check to ensure that truck trips are excluded from counts when comparing to passenger trips from the model • Make sure X-X trips are not double counted • Recheck control totals estimated from the original counts
2. Total implied trip rates including both internal and external travel significantly higher or lower for zones near modeled region boundary	• Check to ensure that trips are not double counted – trips are modeled as external or internal, not both
3. Zones near modeled region boundary have lower percentage of external trips than interior zones	• Check sensitivity of trip generation model to distance from the modeled region boundary

■ 5.3 Commercial Vehicle and Freight Travel

There are a variety of ways in which freight, trucks, and other commercial vehicles are considered in travel models. These include:

- Commodity flow models, where freight movements are estimated by commodity type and converted to truck vehicle trips (and perhaps trips by other freight modes);
- Truck vehicle trip models, usually consisting of truck trip generation and distribution models;
- Factoring of nonhome-based trips; and
- Not explicitly considering truck travel separately.

The latter two methods, of course, are not based on actual data on truck or freight movements and are not recommended. They may be found in many older models or where no local data are available.

Commodity flow models are often found in connection with models of larger regions such as statewide models. Freight flows often cover long distances, and modeling freight for an urban area would mean that a large percentage of the freight modeled is external to the region.

In commodity flow models, the tonnage of freight (either by all modes or only by truck) is estimated using input variables of employment by type and households. The employment types for these models are often more disaggregate than those used in person trip attraction models. These models are often linear regression models, similar in form to person trip attraction models. Normally, one set of regression equations for the production end and one set of regression equations for consumption end are estimated. These regression equations are either developed for each commodity group or type. The outputs of this tonnage generation process are tons of each commodity produced and consumed for each zone.

Following tonnage generation, a tonnage distribution model is run, often a gravity model similar to those used for person trip distribution (see Chapter 6). If multimodal freight tonnage has been generated, a mode choice step follows, where the tonnage tables of freight that are the outputs of tonnage distribution are split into tons by mode (truck, rail, water, air, etc.). Finally, a process to convert tons of truck freight to truck vehicle trips is applied, resulting in truck vehicle trip tables that may be used in the highway assignment process. Truck trips are often modeled by vehicle size category. An example categorization is:

- Light trucks (four-tire);
- Medium trucks (single unit/6+ tire); and
- Heavy trucks (combinations).

It should be noted that there are many trucks that are not carrying freight, and it may be necessary to add these to the freight-carrying truck trips generated by a commodity flow model.

Truck vehicle trip models use truck trip generation equations, which are also usually linear regression models with employment by type and number of households as the independent variables. Truck trip distribution models usually use the gravity model formulation. Truck vehicle trip models are also often segmented by vehicle size category.

There are several other types of commercial vehicles that may be present in modeled regions. These types of commercial vehicles are seldom modeled explicitly although a few models, mainly in large urban areas, do specifically consider taxi trips. A good source for information on commercial vehicle travel is Cambridge Systematics et al, 2004.[16]

5.3.1 Sources of Data

The main sources of data for truck/freight model validation include the following:

- **Local truck/commercial vehicle survey –** If such a survey is available, it is likely to have been the data source for truck/freight model estimation. It is the best source for information on local origin-destination travel data for truck trips. However, such surveys are relatively uncommon.

- **Public/commercial freight data sources –** These include such data sources as the Commodity Flow Survey (CFS), the Freight Analysis Framework (FAF), and TRANSEARCH.

- **Commodity Flow Survey –** The CFS is undertaken as part of the Economic Census through a partnership between the U.S. Census Bureau, U.S. Department of Commerce, and the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation. The survey is undertaken approximately every five years, most recently in 2002 and produces data on the movement of goods in the United States. It provides information on commodities shipped, their value, weight, and mode of transportation, as well as the origin and destination of shipments of manufacturing, mining, wholesale, and select retail establishments. The CFS is available on a CD from BTS.

- **Freight Analysis Framework –** The FAF, available from the FHWA, integrates data from a variety of sources to estimate commodity flows and related freight transportation activity among states, regions, and major international gateways. The original version, FAF1, provides estimates for 1998 and forecasts for 2010 and 2020. The more recent version, FAF2, provides estimates for 2002 and the most recent year plus forecasts through 2035.

  – TRANSEARCH is a freight database that is available commercially from Global Insight. The databases had previously been available from Reebie Associates before they were acquired by Global Insight, and the database is often referred to as "Reebie" data. TRANSEARCH uses a multitude of mode-specific data sources to create a picture of the nation's freight traffic flows on a market-to-market commodity basis. The national database uses counties as the primary flow unit although TRANSEARCH can also use proprietary data to provide a more

---

[16] Cambridge Systematics, Inc., A. Chatterjee, and H. Cohenm, *Accounting for Commercial Vehicles in Urban Transportation Models*, prepared for the Federal Highway Administration, Washington, D.C., 2004.

disaggregate level of geography. Each record in the TRANSEARCH database records the flow from an origin zone to a destination zone.

- **Other national sources –** National data sources include the *Quick Response Freight Manual II*[17] and its predecessor report the *Quick Response Freight Manual*.[18] These reports provide parameter values from other truck/freight models as well as information on other data sources.

- **Vehicle classification counts –** Vehicle classification counts that specify the numbers of trucks by vehicle size category used in the model are critical data needed for validation of truck travel models.

5.3.2 Aggregate Checks

If commercial vehicle or freight models have been developed from local survey data, aggregate checks analogous to those specified for trip generation models might be possible. For example, estimated total truck trips from the model could be compared to truck trips from the expanded truck survey.

Because of the relative lack of local data on truck and freight movements, the main check of truck/freight models cannot really occur until the highway assignment stage, when modeled truck volumes can be compared to vehicle classification traffic counts. If modeled truck volumes are too high (or low) compared to observations, this likely means that the trip/tonnage generation rates are too high (or low), or the truck trip lengths are too long (or short). Vehicle classification counts cannot distinguish full versus empty trucks, freight versus nonfreight related commercial vehicle movements, or intraregional versus interregional truck movements. Thus, the vagaries associated with vehicle classification traffic counts make validation of truck trip generation models after traffic assignment very coarse.

Truck and freight movements can vary significantly depending on geographic, economic, and land use characteristics of an area. It is therefore difficult to determine typical truck travel parameters for areas based on national data sources. The average values provided in these sources may differ significantly from those found in individual areas.

---

[17] Beagan, D., M. Fischer, and A. Kuppam, *Quick Response Freight Manual II*, Publication No. FHWA-HOP-08-010, prepared by Cambridge Systematics, Inc., for the Federal Highway Administration, September 2007.

[18] *Quick Response Freight Manual*, prepared by Cambridge Systematics, Inc., Comsis Corporation, and the University of Wisconsin-Milwaukee, for the Federal Highway Administration, September 1996.

5.3.3 Disaggregate Checks

There are no applicable disaggregate checks of truck/freight models.

5.3.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of truck/freight models.

5.3.5 Reasonableness and Sensitivity Testing

If truck/freight models have been developed using locally collected survey data, the resulting models can be compared to national sources identified in Section 5.3.1. Otherwise, the primary reasonableness checks for truck/freight models are the comparisons to vehicle classification traffic counts, as discussed in Section 5.3.2.

5.3.6 Troubleshooting Strategies

As implied in Section 5.3.2, the calibration strategies for truck/freight models are adjustments to truck model parameters to address significant differences between modeled truck volumes and vehicle classification traffic counts.

5.3.7 Forecasting Checks

Forecast truck/freight travel should be compared to base year for reasonableness. For example, overall numbers of truck trips per employee in the modeled region might be compared for forecast and based years. Also, after the traffic assignment step, the regional truck VMT to total VMT for the modeled region should be compared for reasonableness.

■ 5.4 Special Generators

The term "special generators" refers to specific locations where the trip generation models used for resident internal-internal travel do not accurately estimate the amount of travel generated. These are typically locations that generate high travel demand where typical measures of trip attraction, such as employment and households, are low compared to the demand generated. Typical special generators may include:

- Airports;
- Universities and colleges;

- Recreational attractions;
- Casinos;
- Military bases and establishments;
- Hospitals; and
- Regional shopping centers/malls (usually in smaller modeled regions with few of these locations).

Typically, the estimated demand for each special generator is asserted outside the trip generation process. The demand estimates may be developed from separate person or vehicle counts or from surveys or from data from similar locations within or outside the modeled region. Once the trip ends are estimated, the trips generated by special generators are considered trip productions and attractions for the zones in which the special generators are located, and the productions and attractions are balanced (in four-step models). These trips are then considered through the remainder of the modeling process.

Many special generators are atypical not only in the amount of travel demand generated relative to variables typically used in trip generation models, but also in their trip distribution, mode choice, and time-of-day characteristics. In such cases, it is preferable to conduct surveys of travelers to the special generators to develop separate models for these components. In effect, the special generators are treated as separate trip purposes. For example, many models have reduced the need for special generators by modeling home-based university and airport trips as separate trip purposes. However, it is often infeasible to collect such survey data, due to the expense or the difficulty in obtaining permission to survey at privately owned facilities.

5.4.1 Sources of Data

The main sources of data for special generator model estimation and validation include:

- Special generator traveler survey, as described above; and
- **Traffic or person trip counts** at or near the special generator location.

5.4.2 Aggregate Checks

If counts of trips to/from the special generator are available, the trips generated by the model can be compared to these counts. If such counts are unavailable, it may be possible to conduct traffic counts on nearby roadways leading or adjacent to the special generator, and highway assignment results on these roadways can be compared to the counts.

5.4.3 Disaggregate Checks

There are no applicable disaggregate checks of special generator models.

5.4.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of special generator models.

5.4.5 Reasonableness and Sensitivity Testing

The reasonableness checks for special generator travel models are the comparisons to traffic or person trip counts, as discussed in Section 5.4.2.

5.4.6 Troubleshooting Strategies

Since the demand levels for special generators are asserted, it is a relatively simple matter to adjust these demand levels to address any significant discrepancies between the modeled trips/volumes and the count data described in Sections 5.4.1 and 5.4.2.

5.4.7 Forecasting Checks

Forecast special generator trips should be compared to base year special generator trips for reasonableness. It is important to stratify special generator trips into those expected to increase over time versus those that are expected to remain relatively static. For example, special generator trips for universities and colleges, military establishments, and hospitals might be expected to remain relatively static unless enrollment, troop levels, or numbers of beds are forecasted to change. Conversely, special generator trips to recreational attractions such as parks or open space or to regional shopping centers might be expected to grow with the population of the region even if the basis for the special generator remains unchanged.

6.0 Trip Distribution/ Destination Choice/ Location Choice

6.0 Trip Distribution/Destination Choice/Location Choice

This chapter discusses the validation of model components related to destination choices, or the trip distribution step of four-step models. In activity-based models, relevant model components include regular workplace and school location models, tour-level (primary activity) destination choice models, and trip-level (intermediate stop) destination choice models. All of these types are referred to in this chapter as "trip distribution models."

The most common format for trip distribution in four-step models is the gravity model, an aggregate model structure that estimates a production-attraction trip table from zone-level estimates of trip productions and attractions and measures of separation between zones. Separate models are developed for each trip purpose. In doubly constrained models, the model attempts to preserve the zonal input totals for both productions and attractions; in singly constrained models, it attempts to preserve the zonal input totals for productions only. "K-factors" are sometimes applied to improve the match between modeled and observed trip distribution patterns. Most often, K-factors are applied at a district level, where a district represents a subset of the zones within a modeled region.

In some four-step models, a multinomial logit destination choice model is used. The inputs include the same data as the gravity model, but other inputs may also be used, including socioeconomic characteristics of travelers or households and characteristics of production or attraction zones. A "size variable" corresponding to the attraction zone, generally the number of attractions or another measure of activity (such as a linear combination of employment by type and population or households) is used, usually in logarithmic form.

Multinomial logit is also the most common form of regular workplace location choice, school location choice, and primary activity destination choice in activity-based models. Logit destination choice models are singly constrained.

In activity-based models, or models with components that consider trip chaining, destination choice models for intermediate stops are used. These are also multinomial logit models that use size variables, impedance measures, and potentially characteristics of the trip maker or production/attraction zones. The main difference is that a "detour" impedance measure is used, which is the additional impedance to stop at the attraction zone compared to the impedance to travel directly between the origin and primary activity location.

This chapter is organized as follows: First, checks of the friction factors parameters that define the measure of separation in the gravity model (which is typically used for trip distribution in four-step models) are discussed in Section 6.1.[19] Section 6.2 presents the validation of the trip distribution model along with troubleshooting strategies that may be applied in the model estimation or calibration steps if model validation does not meet desired standards. Checks applicable to both four-step and activity-based trip distribution models are described in Section 6.2.

6.1 Friction Factor Checks

This section discusses checks of friction factors used as parameters in gravity trip distribution models. In logit destination choice models, impedance variables are used directly.

Friction factors are used in the gravity model to represent the effects of travel impedance. Friction factors may be asserted for individual increments of impedance or computed from as a function of travel impedance, usually travel time. The most commonly used friction factor formula is the gamma function:

$$F = a \times t^{b} \times e^{ct} \tag{6.1}$$

Where:

F = Friction factor;

t = travel impedance (usually time in minutes);

a, b, c = model parameters; and

e = base of natural logarithms.

The gamma function reduces to two other commonly used functions for friction factors when either the b or c parameter is estimated to be zero. Specifically, the function reduces to the exponential function when the b parameter is zero, and to the power function when the c parameter is zero. As will be discussed below, both the b and c parameters should typically be negative (if nonzero), producing either a negative exponential function when the b parameter is zero or the inverse power function when the c parameter is zero.

[19] Friction factors define the measure of separation based on travel impedances between TAZs. Checks of travel impedance skims for travel time, travel distance, and travel cost are discussed in Section 3.3. Those basic travel impedance skims are used for both trip distribution and mode choice modeling.

Calibrated or asserted friction factors for individual increments of impedance not fitting a specific mathematical function may also be used. Such friction factors are individually calibrated so that a modeled trip length frequency distribution matches an observed trip length frequency distribution.

6.1.1 Sources of Data

The network data used for developing impedance measures are discussed in Section 3.2. If friction factors are not computed by a formula, they are usually derived from a household travel survey data set or transferred from another model. Some modeling software packages will estimate friction factors by trip purpose directly from a survey data set using the trip length frequency distribution.

6.1.2 Aggregate Checks

Friction factors should be monotonically decreasing from a peak that occurs at a very short travel time (e.g., five minutes or less) or its equivalent in nontime units. This assertion is based on the notion that, everything else being equal, travelers will try to reduce their travel time – traveling five minutes is more desirable than traveling six minutes if the need for travel can be satisfied at either time interval. If a formula such as the gamma function is used to compute the friction factors, this pattern is ensured if reasonable values for the parameters are chosen. The gamma function will be monotonically decreasing if both the b and c parameters are negative. If friction factors are derived directly from survey data, they should be checked to ensure monotonic decreases as travel times increase. While travel survey patterns usually show decreased trip frequency with high impedances, there may be places where the pattern is not "smooth." Care should also be used to ensure that the minimum value for a friction factor is greater than zero; zero friction factors preclude any travel for the impedance range represented.

Periodically, model estimation might result in a positive value for either the b or c parameter. This result might be acceptable in some cases. For example, if the trip distribution model distributes all person trips in motorized vehicles, the function might increase for very small impedance values, such as less than five minutes of travel time, and then monotonically decrease. In other cases, the model might monotonically decrease until an inflection point at a very large value of the impedance (see, for example, the nonhome-based function for "Large MPO 3" in Table 6.1). If that value is substantially larger than would ever be expected in the applications of the model for the region, the positive value for the parameter might be acceptable.

Parameters of formulas such as the gamma function may be checked for reasonableness by comparing them to those used in other models. Table 6.1 presents the "b" and "c" parameters used by six MPOs of varying sizes. The parameters show a relatively large variation. Since friction factors can be scaled without impacting the distribution, the parameters shown in Table 6.1 were scaled to be 1,000,000 at one minute of travel time.

The resulting friction factor curves for the home-based work, home-based nonwork, and nonhome-based trip purposes are shown in Figures 6.1 through 6.3.

Table 6.1 Trip Distribution Gamma Function Parameters

	Home-Based Work		Home-Based Nonwork		Nonhome-Based	
	"b"	"c"	"b"	"c"	"b"	"c"
Large MPO 1	-0.503	-0.078	-3.993	-0.019	-3.345	-0.003
Large MPO 2	-1.65	-0.0398	-1.51	-0.18	-1.94	-0.116
Large MPO 3	-0.156	-0.045	-1.646	-0.07	-2.824	0.033
Medium MPO 1	-0.81203	-0.03715	-1.95417	-0.03135	-1.92283	-0.02228
Medium MPO 2	-0.388	-0.117	-2.1	-0.075	-1.8	-0.16
Small MPO 1	-0.265	-0.04	-1.017	-0.079	-0.791	-0.195

"Large MPO" population is greater than 1 million.

"Medium MPO" population is between 200,000 and 1 million.

"Small MPO" population is less than 200,000.

Note that the gamma function parameters shown in these examples were calibrated based on good approximations of travel times, considering observed levels of congestion, in the various urban areas. Generally, in the larger, more congested areas, this requires feedback of travel times from the highway assignment step. In the smaller areas where congestion is infrequent, the free flow times are reasonable approximations of the travel times experienced on the highway system, and feedback may not be used.

These parameters and graphs are provided as examples and are not intended to provide recommended or transferable friction factors. Differences in urban area geography and other characteristics make it difficult to determine "correct" friction factors for any particular context.

Figure 6.1 Example Home-Based Work Friction Factors Based on Gamma Function

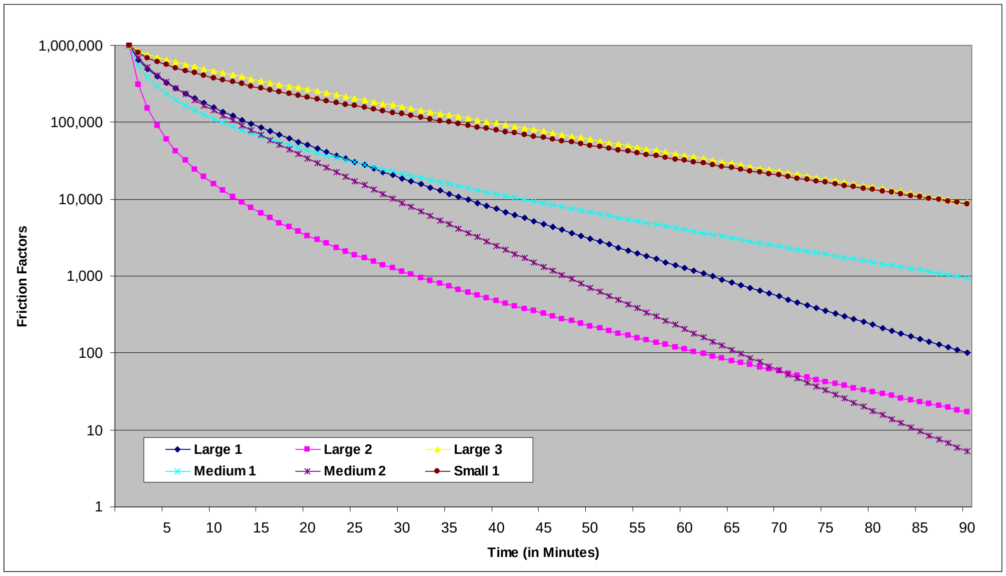


Figure 6.2 Example Home-Based Nonwork Friction Factors Based on Gamma Function

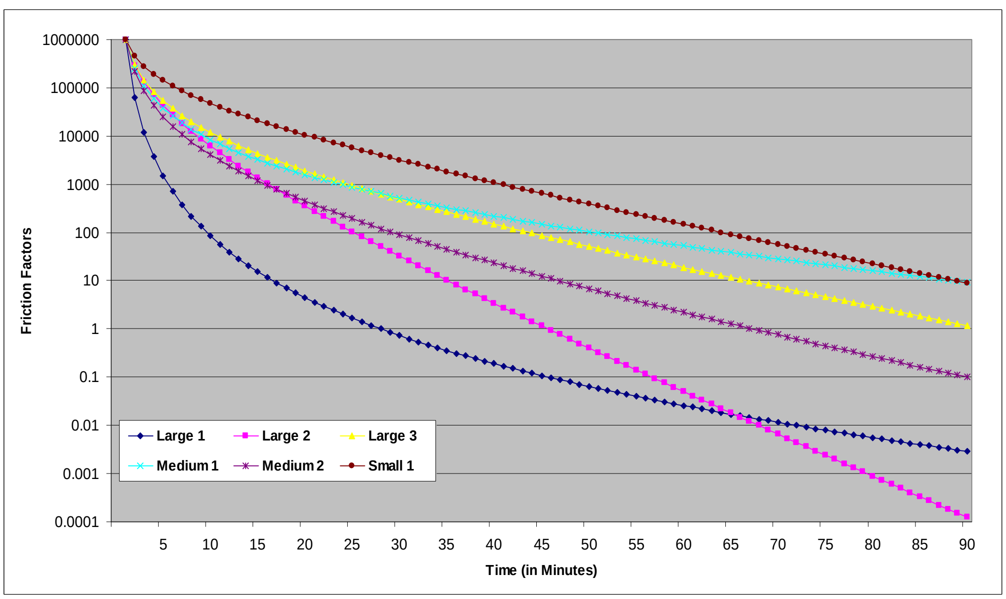
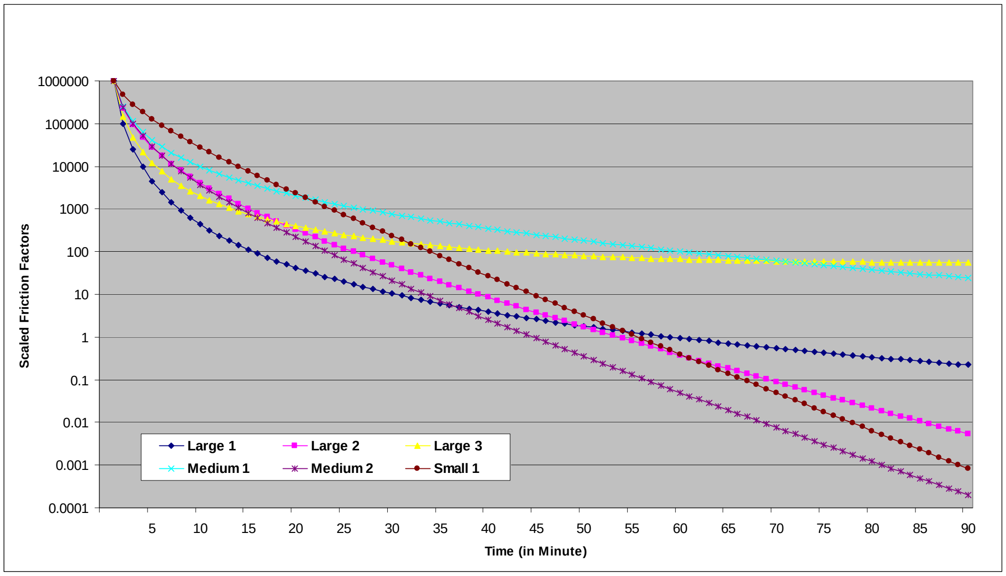


Figure 6.3 Example Nonhome-Based Friction Factors Based on Gamma Function



6.1.3 Disaggregate Checks

There are no applicable disaggregate checks friction factors.

6.1.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of friction factors.

6.1.5 Reasonableness and Sensitivity Testing

Reasonableness checks have been presented as part of the aggregate checks presented in Section 6.1.2. There are no applicable sensitivity checks of friction factors.

6.1.6 Troubleshooting Strategies

Table 6.2 shows some of the potential troubleshooting strategies for dealing with issues with friction factors.

Table 6.2 Troubleshooting Strategies for Issues with Friction Factors

Issue	Potential Troubleshooting Strategies
1. Friction factors do not monotonically decrease from a peak occurring at a very short travel time	• Check parameters of friction factor formula (e.g., gamma function) • Check friction factor pattern for match with observed trip length frequency • Check to ensure that "smoothed" observed data pattern is monotonically decreasing

6.2 Checks of Trip Distribution Model Results

Generally speaking, the types of models described in Chapter 6 match production, origin, or home zones with the locations of trip attractions, destinations, or activities. For the trip distribution component of four-step models, these results are presented as trip tables. For components of activity-based models, the activity locations are modeled relative to the home location (or the work activity location in the case of a work-based subtour), but trip tables are not created; the results of other model components (e.g., daily activity patterns, intermediate stop models) are used so that trip tables are created later if needed. The outputs in these cases are matched origin and activity/destination locations. In the case of intermediate stop location models, the activity location is determined based on the locations of both a home/origin and a primary activity location. For simplicity, the results to be validated will be referred to as "trip tables" throughout this chapter.

6.2.1 Sources of Data

The main sources of data for validation include the following:

- **Household travel/activity survey –** If such a survey is available, it is likely to have been the data source for model estimation. It is the best source for information on local origin-destination travel data for all person trips. It can be used to produce trip length frequency distributions and can be expanded to produce a trip table of all travel made by the residents of a modeled region. The expanded trip table will not be statistically valid at the zone level but can be used at a more aggregate (district) level.

- **Census data –** Through the 2000 U.S. Census, the long form provided the information for the Census Transportation Planning Package (CTPP). The long form asked for information on the journey to work, including work location, departure time, and primary mode used. Since the long form has been eliminated for the 2010 Census, the Census Bureau has developed the American Community Survey (ACS), which is conducted continuously, not only at the time of the decennial census. The ACS also collects data on work location and travel.

Work travel is treated differently in the ACS compared to travel models. The ACS asks:

– Where each person worked "most last week";

– The street address of the workplace;

– How the person "usually" traveled to work (single mode "used for most of the distance");

– The "usual" auto occupancy;

– The "usual" departure time from home;

– The usual travel time in minutes; and

– Whether the person was temporarily absent from work due to layoff, vacation, illness, labor dispute, etc.

Note that the responses to these questions differ from the way work travel may be treated in household surveys and models in several ways:

– A household survey asks about the respondent's travel to work on the travel day, as opposed to where he or she usually worked over the past week;

– Mode is defined in more detail in travel surveys and models (e.g., park-and-ride as opposed to bus or auto);

– A household survey asks about stops on the way to or from work, whereas the ACS does not ask about stops on the way to work; and

– Home-based work trips in models include trips home from work as well as to work.

- **National sources** – National data sources include the National Household Travel Survey (NHTS), NCHRP Report 365, Travel Estimation Techniques for Urban Planning, which is being updated (the update is expected to be available in 2010), and other documents (e.g., TCRP Report 73, Characteristics of Urban Travel Demand).

6.2.2 Aggregate Checks

There are two general types of aggregate checks of trip distribution model results: Trip length checks and origin-destination pattern checks, which are associated with base year scenarios. Scenario comparisons are an aggregate check appropriate for forecast and other types of alternative scenarios for which observed data may not be available and are discussed in Section 6.2.5.

Trip Length Checks

The main trip length checks for base year scenarios involve comparing average trip lengths and trip length frequency distributions between model results and observed data from the household travel survey. For home-based work trips, CTPP data are often a supplementary source for trip length information. Trip length checks should be performed separately for each trip, tour, or activity purpose – essentially for each separately estimated model.

Trip lengths should be checked as represented in both distance and time units. While the impedance measure may consist of other level of service elements, it is important to check (and easier to interpret) results for time and distance. However, "trip length" checks may be performed for the specific measure of impedance used in the model as a supplement to the time and distance checks.

Most travel surveys ask respondents to report departure and arrival times and therefore provide estimates of trip lengths. However, experience with trip length information from these surveys indicates that they do not provide reliable information for developing average trip lengths or trip length frequency distributions. It is easier for respondents to recall and report specific activities and locations than to do so for specific departure and arrival times. Even when the times are recalled accurately, many respondents round times to the nearest 5, 15, or 30 minutes. Travel times can vary significantly from day to day, and models use travel times based on average conditions, which are closer to what travelers assume in decision-making.

To address this issue, trip lengths are computed using the time and distance skims applied to the specific origins and destinations reported in the survey. Average trip lengths and trip length frequency distributions for the observed condition are computed directly from the trip table obtained from the expanded survey data and compared to trip table information obtained from applying the model. For the modeled trip tables, trip generation estimates must be available since they are necessary inputs to the trip distribution model (the same applies to the models providing inputs to activity-based location choice models).

The first check is to compare modeled to observed average trip lengths. These should be checked separately for each trip or activity purpose. It is also desirable to check trip lengths by market segment, with segments defined however possible given the model's capabilities and the information available from the observed survey data. For example, if trips by different income levels are modeled separately, it would make sense to compare average trip lengths for each income level modeled. Another possible segmentation scheme is by geographic subregion or district. For a disaggregate activity-based model, market segments can be defined in numerous ways, limited only by the variables used in the definition of the synthetic population and the variables included in the survey.

It is insufficient to check only the average trip lengths; the frequency distribution of trip lengths must also be checked. Visual checks can be very useful; the observed and modeled trip length frequency distributions can be plotted on the same graph to see how closely the distributions match.

A more rigorous way of checking trip length frequency distributions is through the use of *coincidence ratios*. This is most easily understood as the area under both curves divided by the area under at least one of the curves, when the observed and modeled trip length frequency distributions are plotted. Mathematically, the sum of the lower value of the two distributions at each increment of time or distance is divided by the sum of the higher value of the two distributions at each increment. Generally, the coincidence ratio measures the percent of area that "coincides" for the two curves.

The procedure to calculate the coincidence of distributions is as follows:

$$CR = \left\{\frac{\sum_T[min(PM_T, PO_T)]}{\sum_T[max(PM_T, PO_T)]}\right\}$$

where:

CR = Coincidence Ratio

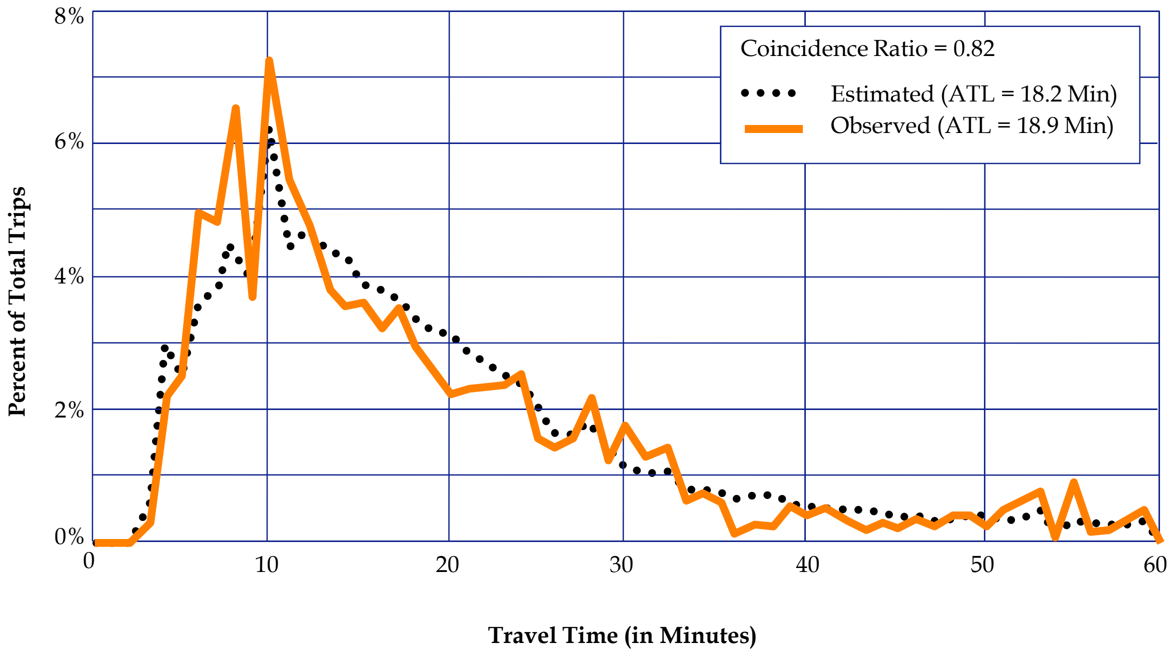
PM_T = Proportion of modeled distribution in interval T

PO_T = Proportion of observed distribution in interval T

T = Histogram interval for time, distance, or other impedance measure (e.g., 0-4.9 minutes, 5.0-9.9 minutes…)

The coincidence ratio lies between 0 and 1.0, where a ratio of 1.0 indicates identical distributions. Figure 6.4 shows an example application of the coincidence ratio.

Figure 6.4 Coincidence Ratio



Note: ATL = Average Trip Length.

Origin-Destination Pattern Checks

As discussed previously, it is impossible to estimate a statistically valid origin-destination trip table at the zone interchange level by expanding household surveys with typical sample sizes. It is therefore necessary to check origin-destination patterns at a more aggregate level. Generally, this is described as a *district*-level validation. The ideal number of districts is dependent on many factors, including the size of the modeled region, the number of zones, the amount of travel, the existence of political boundaries and travel barriers such as rivers, and the amount of market segmentation for which district-level analysis will be performed.

In most urban area models, the central business district is analyzed as a separate district or group of districts. It is also good practice to keep major political entities (states, counties, and large cities) in separate districts wherever possible. Many types of trips (e.g., school trips) can be affected by these political boundaries. Significant travel barriers, including major rivers or rail corridors with limited crossing opportunities, make good district boundaries. It may also make sense to align district boundaries with screenline or cutline locations (see Section 9.1) to better use information from highway assignment in validation of travel patterns.

As with other checks, district-level geographic checks should be performed separately for each trip or activity purpose. Additional market segmentation, such as by income level, should also be performed where the observed data exist and the model supports such segmentation.

Another check of origin-destination patterns is the *orientation ratio*. This ratio measures the propensity of trips from a production area (or zone) to the attraction area and is defined as follows:

- The numerator is the trips to the attraction area from the production area divided by the sum of all trips to the attraction area; and
- The denominator is all trips from the production area divided by all trips in the region.

For example, say that from district A to district B, there are 300 trips, and there are 800 total trips produced in district A, 100,000 total trips attracted to district B, and 750,000 total regional trips. The orientation ratio for district A to district B is computed as:

(300/100,000) / (800/750,000) = 2.81

An orientation ratio with a value of less than 1.0 indicates that the production area is less oriented to the attraction district than is the region as a whole. An example might be a low-income residential area located adjacent to a “tech center” area. Conversely, an orientation ratio with a value of greater than 1.0 indicates that the zone is more orientated to the attraction area than other zones in the region. The orientation ratio can be calculated on a zonal level or a mixed zonal and district level, as desired.

The orientation ratio can be useful for understanding the reasonableness of modeled distributions. If modeled distributions for selected areas are compared to observed distributions for validation purposes, the checks should be performed on a district basis due to the inherent sampling error associated with household surveys. As a district-level validation check, the observed and modeled orientation ratios can be calculated for each district-to-district interchange. Then, the ratios of the modeled to observed district-level orientation ratios can be estimated. The closer the ratios are to 1.0, the better the model is reproducing observed orientations. Likewise, the differences between the modeled and observed orientation ratios can be checked. Differences close to zero imply better reproduction of observed trip orientations.

Plots of orientation ratios can also be used to qualitatively check trip distribution results. Figure 6.5 shows an example of modeled and observed orientation ratios estimated on a zonal basis (for the productions) to an attraction district (the CBD). The calculated ratios have been plotted. The results shown in Figure 6.5 suggest that there might be too much modeled travel from the eastern part of the region to the CBD. The observed results in Figure 6.5 also show the difficulty with using travel survey data on a zonal level. Aggregation of the modeled and observed results shown in Figure 6.5 to a district-level structure might provide a clearer picture of differences in the modeled and observed orientations of trips to the CBD for this example.

Another geographic check of travel patterns is the amount of intrazonal travel. This depends greatly on the definition of zones and the level of detail of the model. The modeled percentage of intrazonal trips should be compared to the observed percentage of intrazonal travel from the expanded household survey. These checks should be done separately by trip purpose and market segment and may also be done at the district level.[20]

6.2.3 Disaggregate Checks

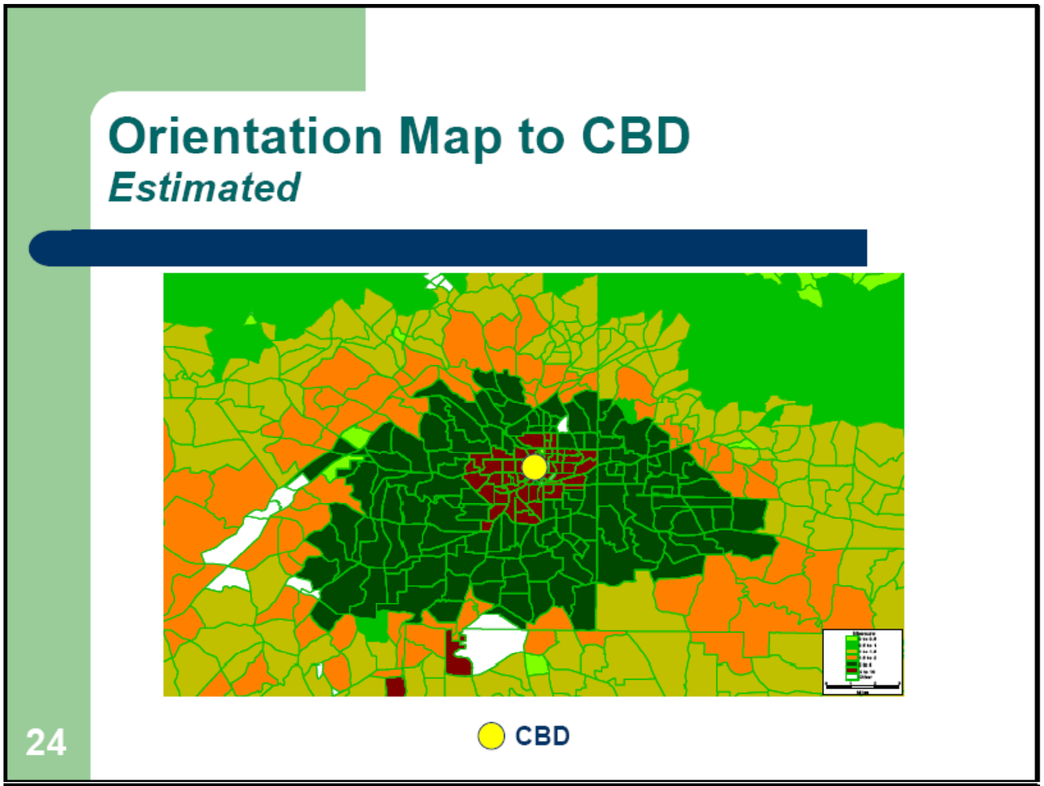
Since the gravity model is an aggregate formulation, there are no applicable disaggregate checks. Logit models are disaggregately estimated (one record per trip/activity), and therefore disaggregate validation is needed for logit destination choice models, along with the aggregate checks described above.

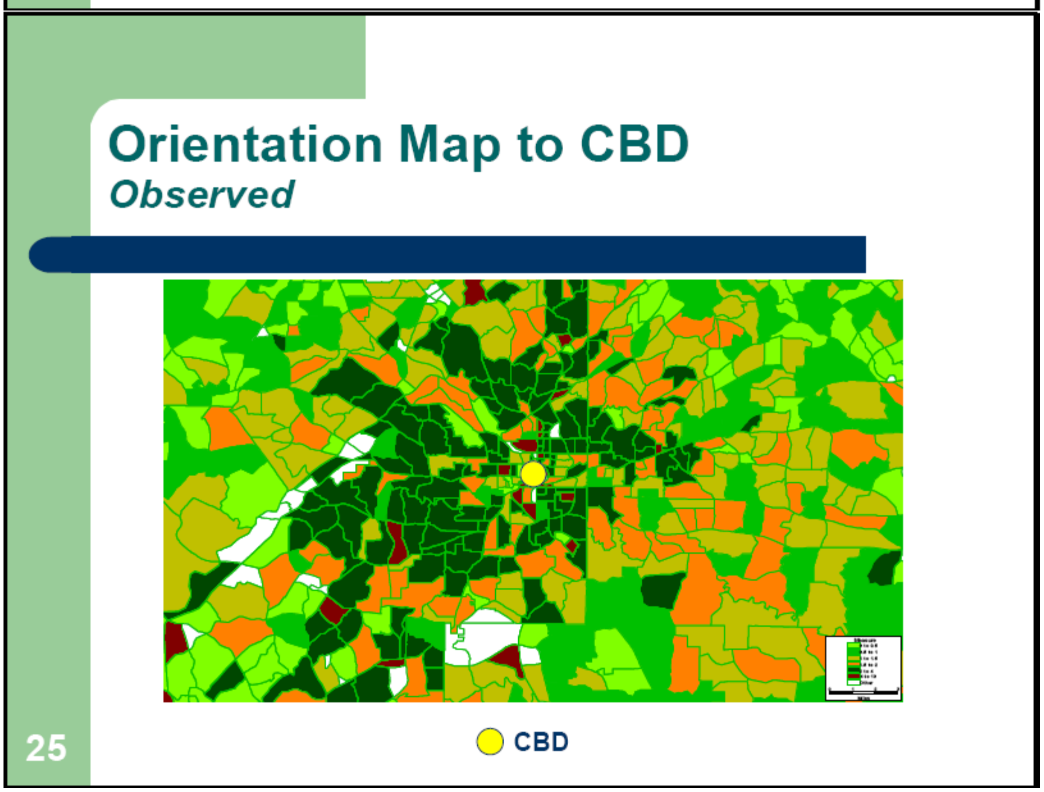
Ideally, disaggregate validation of a model should be performed using a data set that is independent of the data set used for model estimation. For example, a survey data set might be divided in half, with one half used for model estimation and the other for model validation. As a practical matter, however, most urban area household travel surveys have such small sample sizes that the entire data set is needed for model estimation. In most cases, therefore, there is no independent model estimation data set available for validation.

Limited disaggregate validation can be performed using the same data set used for model estimation, but reporting the results by market segment. Logit model estimation software has the capability to apply the estimated model to a data set in the same form as the estimation data set. For example, a logit destination choice model could be applied to the data set used for estimation but the results may be reported by vehicle availability level. It might be found, for example, that zones in the CBD are not being chosen often enough in the model for households with zero vehicles.

[20] Zone-level intrazonal percentages from the survey data set are not statistically significant in most cases.

Figure 6.5 Example Orientation Ratio GIS Maps





Source: "*Shining a Light Inside the Black Box,*" TMIP webinar, March 11, 2008.
Note: This figure is not critical in understanding the concept of orientation maps.

While it is good practice to perform disaggregate validation of logit destination choice models, it is often difficult to use the results. Since a destination choice model usually consists of zones, there are hundreds or thousands of alternatives, making the reports of the model application very long and time consuming to analyze. Furthermore, as discussed above, typically household survey sample sizes do not have large enough sample sizes to analyze results at the zone level. While disaggregate validation is a useful tool to identify systematic biases in the estimated models, there are often aggregate tests that are more practical. For example, model application results could be segmented by trip length category, but the aggregate trip length checks described in Section 6.2.2 provide a much easier way of identifying whether modeled trip lengths are accurate.

6.2.4 Criteria Guidelines

The guidelines presented in this chapter are intended for use with the comparisons of base year modeled results to observed data shown in Section 6.2.2. As with criteria guidelines throughout the validation process, the figures presented in this chapter are guidelines and are not to be interpreted as absolute standards. Meeting every guideline does not ensure a validated model, nor does missing a guideline imply that the model is not validated.

Average trip lengths – Generally, the modeled average trip lengths for each trip purpose should be within five percent of observed for each trip purpose. In models with many trip purposes, some purposes may have relatively few trips, and so the five percent guideline can be relaxed in these cases.

Coincidence ratios – It is preferable for the coincidence ratio for each trip purpose to be at least 70 percent. The 70 percent guideline can be relaxed in models with many trip purposes since some purposes may have relatively few trips.

Intrazonal trip percentages – Some models have used guidelines where regionwide intrazonal trip percentages should be within three percentage points of observed for each trip purpose. For example, if a trip purpose had an observed intrazonal trip percentage of seven percent, the modeled percentage should be between 4 and 10 percent.

Orientation ratios – Orientation ratios are a relatively new measure and, as a result, no guidelines have been set. However, as described in Section 6.2.2, if the ratios of modeled to observed orientation ratios are calculated, values closer to 1.0 imply a better match of the modeled distribution to the observed distribution. Likewise, if the differences between the ratio are calculated, values closer to zero imply a better match of modeled to observed distributions.

6.2.5 Reasonableness and Sensitivity Testing

The reasonableness checks for trip distribution models are mainly the aggregate checks presented in Section 6.2.2.

Sensitivity testing can be performed for trip distribution models by varying model inputs and checking results for reasonableness. Model inputs that can be varied include level of service variables used in the trip distribution model (time/speed and cost) and the level of activity/number of trips in different parts of the region. Some example tests include the following:

- Increasing or decreasing speeds by a fixed percentage regionwide or on a specific subset of highways (e.g., freeways);
- Increasing/decreasing parking costs in the CBD by a fixed percentage;
- Increasing/decreasing tolls by a fixed percentage or amount;
- Increasing/decreasing travel times, headways, or fares by a fixed percentage;
- Changing development patterns for forecast years by moving projected new activity among different parts of the modeled region (e.g., from suburbs to small urban centers or from outlying areas to infill); and
- Reallocating the number of household by income level for a forecast year.

Due to a lack of experience with sensitivity testing for trip distribution models, developing standards for expected changes in trip lengths or distribution patterns based on changes in input variables is difficult. While elasticities of travel in response to changes in transportation level of service are generally well below 1.0 in absolute value, it is up to the experienced planner to determine whether the modeled changes are reasonable.

Scenario Comparison

For scenarios where there are no observed data for comparison, model results must be checked by comparing them to another previously validated scenario such as the validated base year scenario, another forecast scenario, or a "no build" scenario associated with the alternative policy/project scenario.

All of the checks listed previously in Section 6.2.2 can be used for these comparisons, including average trip lengths and trip length frequencies, district-level origin-destination flows, and orientation ratios. Orientation ratios can provide a good reasonableness and sensitivity check. Similar plots to those shown in Figure 6.5 could be produced for trip distributions performed for base and alternative scenarios. Resulting changes in orientation ratios should be visible in a comparison of the plots.

The purpose of these comparisons is not to match the model results with those from the comparison scenario, but rather to determine whether the differences are reasonable. It is therefore important that those performing these checks include persons with knowledge of the modeled region as well as the models themselves.

6.2.6 Troubleshooting Strategies

Issues discovered during the model checks described in Sections 6.2.2, 6.2.3, and 6.2.5 may imply errors in trip distribution model parameters or input data (networks/skims or trip ends). Table 6.3 shows some of the typical problems that may be evident from these tests. Because it is a critical topic, one strategy, the use of K-factors, is discussed in more detail at the end of this chapter.

Table 6.3 Troubleshooting Strategies for Issues with Trip Distribution Model Results

Issue	Potential Troubleshooting Strategies
1. Average trip lengths too long or short	• Recalibrate friction factors or adjust parameters of friction factor formula or logit utility equations • Recheck skim data and trip end inputs • Check distribution patterns (see below)
2. Coincidence ratio too low	• Recalibrate friction factors or adjust parameters of friction factor formula or logit utility equations
3. District-level origin-destination patterns inaccurate for some interchanges	• Check trip lengths (see above) • Check travel impedances between affected districts • Introduce or adjust K-factors • Introduce impedance penalties on network links (e.g., bridge crossings)
4. Too may or few intrazonal trips	• Adjust intrazonal travel times for types of zones with this issue
5. Model too sensitive or insensitive to changes in level of service	• Adjust parameters for appropriate level of service variables in impedance/utility functions or friction factors

Tests described here are not intended to be performed in isolation, and an iterative approach to testing may be necessary. For example, consider an urban area that includes parts of two states. If the model is predicting too much travel between the two states, the model may not be sufficiently sensitive to the reasons why interstate travel might be avoided for some purposes, or it could simply be overestimating longer trips (which may be biased toward interstate trips). It is necessary to check the results of both the district-level origin-destination summaries and the trip length frequency distribution comparison to determine the most likely causes for these model results and to suggest the best calibration methods. The best method could be the introduction of interstate K-factors, changes to the friction factors, or some combination of both of these.

K-Factors

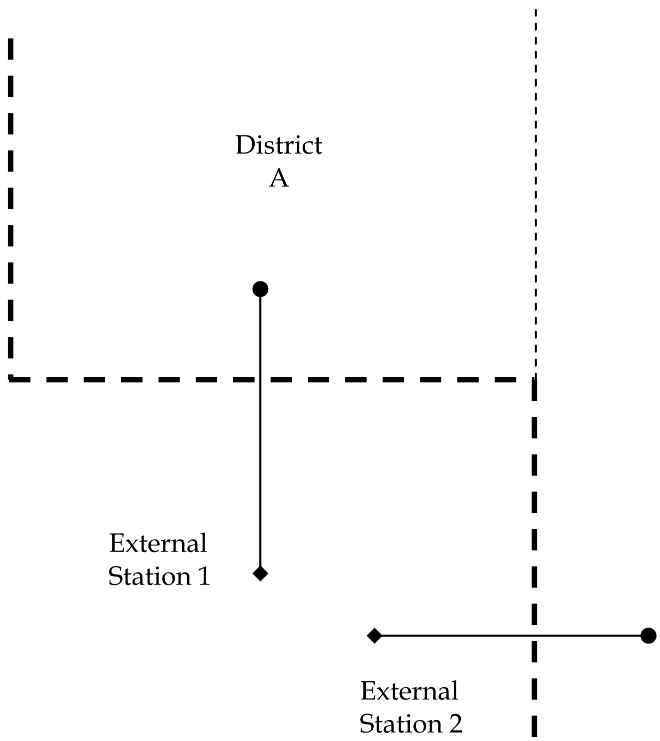
K-factors correct for major discrepancies in trip interchanges, usually at the district level. They are typically justified as representing socioeconomic or other characteristics that affect trip making but are not otherwise represented in the model. Physical barriers, such as a river crossing, may also result in differences between observed and modeled trip patterns. For example, trip movements between zones separated by a bridge may not be as great as would be expected using only quantifiable measures. In this case, a planner could use either K-factors or artificial times on the bridge links to match the actual interchange of travel.

In a sense, K-factors are analogous to the alternative specific constants in logit models; they are intended to account for the choice factors that are not able to be included in the models. Since trip distribution models have relatively few input variables, it is reasonable to believe that other factors that affect location choice are not included in the models. In many cases they cannot be measured, quantified, or forecasted. K-factors provide a means for accounting for these factors, although they are then assumed to remain fixed over time and across all scenarios.

For this reason, K-factors must be used very cautiously. Because they can be used to provide nearly perfect matches between modeled and observed district-level origin-destination flows, it can be very tempting to apply K-factors to resolve differences in origin-destination flows without determining whether they are the best method to solve the problem at hand. The use of K-factors, therefore, should be considered “a last resort” after all other possible causes for error and calibration adjustments have been considered. Even when K-factors are introduced, they should be relatively small in magnitude – the closer to 1.0, the better.

A valid use of K-factors may occur when it is necessary to “zero out” trips between groups of zones. Consider the example of a trip distribution model of external-internal trips for the region depicted in Figure 6.6. Because of the geography of the modeled region boundary, it is unlikely that trips would occur between External Station 2 and the zones in District A. Trips from the area outside the modeled region as shown in Figure 6.6 would likely enter the modeled region via External Station 1 to travel to zones included in District A. However, internal-external trip productions from the zone representing External Station 2 in the travel model would not be precluded from “traveling” to zones in District A in a typical trip distribution process. So, introducing a K-factor of zero between External Station 2 and the zones in District A would ensure that these unlikely trips would not occur in the model.

Figure 6.5 Example of K-Factor Use with External Stations



7.0 Mode Choice/ Vehicle Occupancy

7.0 Mode Choice/ Vehicle Occupancy

This chapter discusses the validation of model components related to mode choices,[21] including estimation of vehicle occupancy levels. In activity-based models, relevant model components include tour- and trip-level mode choice models.

The most common formats for mode choice models are the multinomial and nested logit models. The inputs always include level of service variables such as time, cost, and the number of transit transfers. Other input variables may also be used, including socioeconomic characteristics of travelers or households and characteristics of production or attraction zones. Mode choice is usually modeled separately for each trip or tour purpose.

The alternatives for logit mode choice models are the travel modes. **Auto modes** often include classifications by vehicle occupancy level, with the highest level being determined by analysis needs such as the need to distinguish between two and three person carpools for high occupancy vehicle (HOV) lane analysis. Some models also classify auto submodes by whether or not the path includes paying a toll. A simple model could have only one or two (drive alone versus carpool) auto submodes; a more complex model could have many auto submodes (drive alone-toll, drive alone-no toll, two person carpool-toll, etc.).

Transit modes are usually divided into walk access and auto access submodes if there are a significant number of transit riders who drive to or are dropped off at transit stops. Some models further separate the auto access submodes into park-and-ride and kiss-and-ride (drop-off) submodes. In areas with significant transit service that is not local or line-haul bus, the transit submodes may be further disaggregated into submodes defined by technology or operating characteristics. These submodes may include local bus, express bus, light rail, subway/elevated, commuter rail, and other modes. A more complex model might include many transit submodes defined by access mode and technology/ operation.

Nonmotorized modes include walking and bicycling. In nearly all cases, there is simply a single nonmotorized mode or two distinct modes, representing walk and bicycle.

Some regions might specify more complex travel models than are warranted based on existing travel choices represented in a travel survey. This is often the case for regions

[21] If a model is vehicle-trip based (i.e., if vehicle trips rather than person trips are generated and distributed), there is no need for mode choice or vehicle occupancy modeling capabilities. The information presented in this chapter is unnecessary for validation of such models.

that anticipate testing new alternatives in the future. For example, a region that does not have HOV lanes in a base year might specify a model that stratifies the auto mode by group size rather than simply modeling drive alone and carpool trips. This would provide the capability for testing alternative carpool lane treatments. Likewise, an area without fixed guideway transit service may specify mode models that include walk and drive access (to estimate park-and-ride lot usage) and fixed guideway submodes in order to test future alternatives.

Vehicle occupancy estimates may be obtained from the mode choice model if drive alone and carpool submodes are modeled separately. However, in these cases, it is always necessary to make assumptions about vehicle occupancy for the submode representing the highest occupancy level. For example, if a model has three auto submodes – drive alone, two person carpool, and three or more person carpool – the vehicle occupancy level for the three or more person carpool mode must be assumed. The highest occupancy level, which may vary by trip or tour purpose, is typically estimated from household survey data. If there are not separate auto submodes for different occupancy levels, a separate vehicle occupancy model, usually separated by trip/tour purpose, may be applied to convert auto person trips to vehicle trips. This may be a simple model with inputs based on trip length.

This chapter is organized as follows. Validation of mode choice models is discussed in Section 7.1. Section 7.2 presents the validation of vehicle occupancy levels, whether they are estimated in the mode choice model or through other means.

■ 7.1 Checks of Mode Choice Model Results

The outputs of mode choice models are trip tables, tours, or trips by mode. Since the inputs indicate the origins and destinations of the tours or trips, the geographic locations of the results are known. This allows geographic segmentation of the results for validation.

7.1.1 Sources of Data

The main sources of data for validation of mode choice models include the following:

- **Transit ridership counts –** Transit ridership counts have the best information on the total amount of travel by transit, usually at the route level. It is important to recognize, however, that ridership (boarding) counts represent "unlinked trips," meaning that a person is counted each time he or she boards a new transit vehicle. So a trip that involves transit transfers is counted multiple times. Mode choice models generally consider "linked trips," where a trip including transfers counts as only one trip. Information on transfer rates is required to convert unlinked trips to linked trips; such information generally is obtained from transit on-board surveys.

- **Transit rider survey –** A transit rider survey (typically an on-board survey) is an invaluable source of information for validation of the transit outputs of mode choice models but may have also been a data source for model estimation. A wealth of information that cannot be obtained from transit counts is available from on-board surveys, including:
  - Transit trip origin-destination patterns by trip purpose;
  - Access modes;
  - Transit paths (surveys should ask riders to list all routes used in order for the linked trip);
  - Transit submodes used (e.g., bus, light rail);
  - Transit transfer activity; and
  - Characteristics of the surveyed riders and their households.

  It should be noted that transit on-board surveys usually provide data only for individual transit trips, not tours, and so their use in estimating transit travel in tour-based models is limited.

- **Household travel/activity survey –** If such a survey is available, it may have also been a data source for model estimation although data from other sources such as transit on-board surveys may also have been used in model estimation. The household survey is the best source for information on nontransit travel data since the number of observations for transit travel is usually small. The expanded household survey data can be used to produce observed mode shares for nontransit travel by purpose for a number of geographic and demographic market segments.

- **Census data –** The Census Transportation Planning Package (CTPP) contains information on modes for work travel. As discussed in Section 6.2.1, the Census Bureau now uses the American Community Survey (ACS), which is conducted continuously, to collect data on work location and travel (among other items). Section 6.2.1 also discusses how work travel is treated differently in the ACS compared to travel models. The CTPP data are more consistent with tour-level work travel than trip level; however, data on travel only in the home to work direction are available.

- **National sources –** National data sources include the National Household Travel Survey (NHTS), NCHRP Report 365, Travel Estimation Techniques for Urban Planning, which is being updated (the update is expected to be available in 2010), and other documents (e.g., TCRP Report 73, Characteristics of Urban Travel Demand).

- **Highway usage data –** Data on highway usage such as toll road and high-occupancy vehicle lane use would be helpful validation data for models that include related modal alternatives.

7.1.2 Aggregate Checks

The most basic aggregate checks of mode choice model results are comparisons of modeled trips or tours by mode, or mode shares, to observed data by market segment. Market segments include trip or tour purposes as well as demographic segments, such as income or vehicle availability levels, and geographically defined segments.

Mode choice models are typically applied using trip tables (or their tour-based equivalents) as inputs. The mode choice model's results, therefore, are shares of the total trip table for each market segment that use each of the modal alternatives. Validation of the model's aggregate results involves checking the shares for the model's base year scenario results against observed mode shares.

A household survey is the only comprehensive data source covering all modes, and therefore is the only source for mode shares. However, mode shares for modes that are used relatively infrequently – notably transit modes – as well as mode shares for relatively small segments of the population (such as members of zero-vehicle, high income households) cannot be accurately estimated from household surveys due to small sample sizes.[22] While it may be problematic to find an alternate source for some segments or modes (such as bicycle travel), transit trips and shares by segment may be estimated using data sources including ridership counts and transit rider surveys.

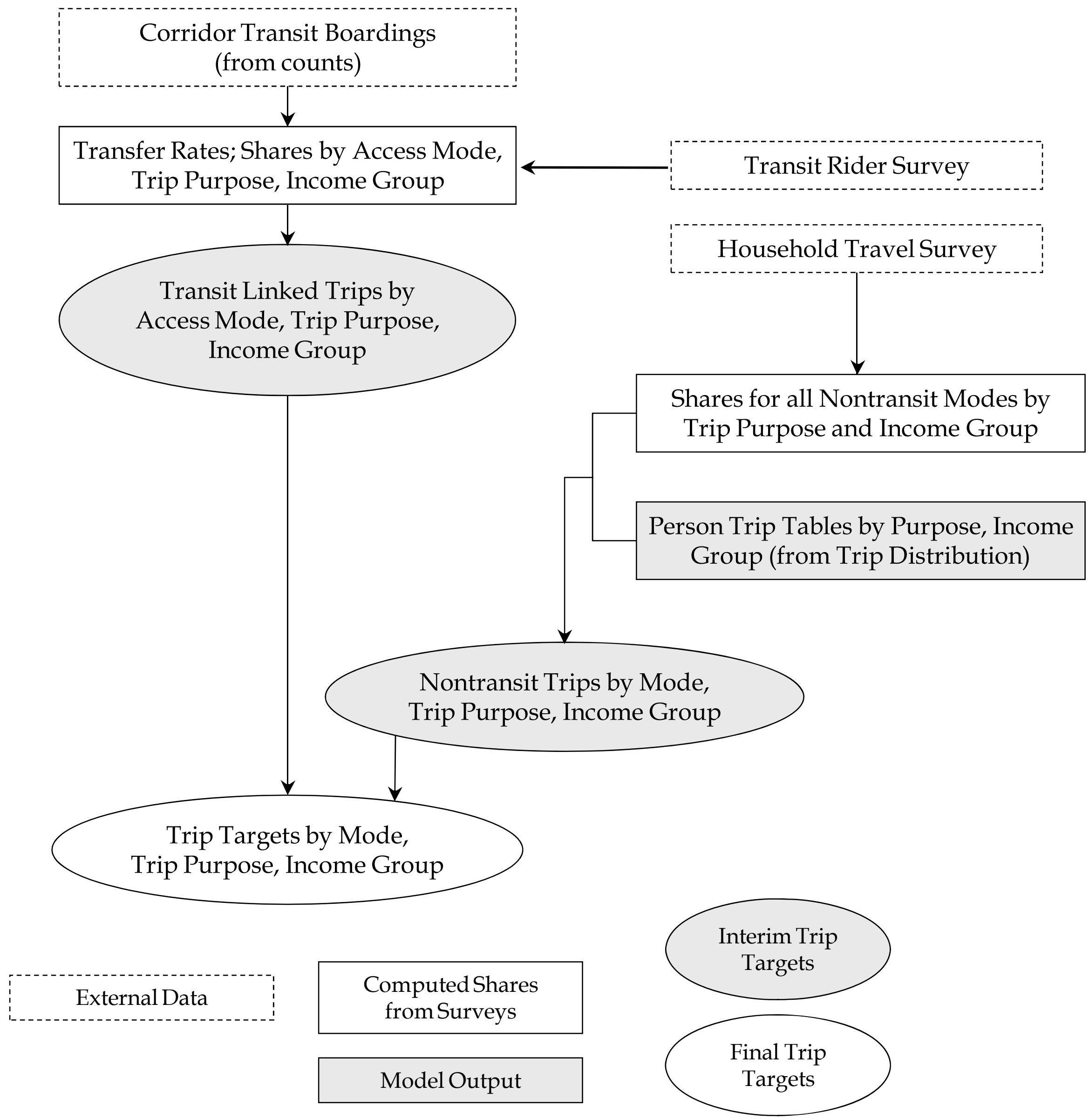
Transit ridership counts provide estimates of total transit *trips*, not mode shares. To convert these trips to shares, an estimate of the total trip table for each market segment is needed. Assuming good validation of the trip generation and distribution components (or their tour-based equivalents), the trip table outputs from the trip distribution model can provide this information. Basically, the transit trips by submode, access mode, trip purpose, and other segmentation level, segmented using the transit rider survey data, can be subtracted from the total trips represented in the trip table to obtain estimates of "observed" nontransit trips. The nontransit trips can be separated into trips by individual mode (auto and nonmotorized submodes) using information from the household travel survey. An example of how this could be done is presented below.

Example for Estimating Observed Travel By Mode

Figure 7.1 shows a method for estimating observed travel by mode. Say that a trip-based mode choice model is to be validated for a market segment defined geographically (a particular corridor), temporally (a.m. peak period) and demographically (the lowest income group). It is assumed that the trip generation and trip distribution steps are segmented by time period and income group, for each trip purpose. Finally, say that the model has four modes: auto, transit-walk access, transit-auto access, and nonmotorized.

[22] Some household surveys are designed to over-sample small segments of the general population in order to address these issues. The surveys must be carefully designed and administered to reduce sample bias.

Figure 7.1 Example of Transit Mode Choice Validation Target Development



Transit-walk access mode – The transit ridership counts for the corridor for the a.m. peak period are obtained. These counts, which represent boardings, are adjusted using information from on-board survey data to represent "linked" trips (where a trip with a transfer is not counted twice), and the linked trips are separated by trip purpose. The share of transit trips that are walk access is estimated from the on-board survey data, as is the share of trips made by the lowest income group. (Note that this need not be a sequential process; the adjustments for transfers, trip purpose, access mode, and income level could be done in a combined way using the survey data.) The result is an estimate of a.m. peak period transit-walk access trips made by low income travelers in the corridor, stratified by trip purpose.

Transit-auto access mode – These trips are estimated in essentially the same way as the transit-walk access trips, except that transit-auto access travel data from the on-board survey are used. The result is an estimate of a.m. peak period transit-auto access trips made by low income travelers in the corridor, stratified by trip purpose.

Nonmotorized trips – The transit trips by purpose are subtracted from the total trips made by low income residents in the a.m. peak period the corridor, which are known from the outputs of trip distribution. The share of nonmotorized trips among nontransit trips for this market segment can be obtained from the household travel survey data. The result is an estimate of a.m. peak period nonmotorized trips made by low income travelers in the corridor, stratified by trip purpose.

Auto trips – The remaining trips in the corridor for this market segment must be by auto. The result is an estimate of a.m. peak period auto trips made by low income travelers in the corridor, stratified by trip purpose.

The results of the mode choice model for the segment can be compared to these estimates of observed travel for aggregate model validation. However, it is also a good idea to estimate results for the other segments – the other income groups for the time period – in the corridor and evaluate all the estimates of observed travel for reasonableness prior to performing the aggregate model checks. It might also be wise to stratify the estimates by interchange distance since nonmotorized trips are likely to vary substantially by distance.

Other Considerations

It is common practice to compare overall mode choice model results to the observed data for the region. It is important to recognize that this type of regional check is not sufficient to determine that the mode choice model is validated, any more than it would be sufficient to validate a highway assignment model simply by comparing total regional vehicle miles traveled (VMT) to observed VMT. Aggregate validation must also be performed for all relevant market segments for which information can be obtained.

The mode choice model validation process is tied in with the transit assignment validation process, which is described in Section 9.2. Any calibration of the transit assignment process may lead to model changes that affect mode choice, whether they are network changes, revisions to path building or skimming, or other changes to the model. The mode choice models cannot be considered validated until the transit assignment model has also been validated.

Transit Trip Lengths

If observed data on transit trip lengths are available, modeled transit trip lengths should be compared to the observed data. While this is a check of both trip distribution and mode choice, the mode choice model must be run before this check can be performed.

Data on transit trip lengths is usually obtained from transit rider surveys. There are two levels at which observed transit trip length data may be available:

- For the in-vehicle portion of transit trips (stop to stop); and
- For entire trips (origin to destination).

Modeled trip lengths can be obtained for either level although the analyst should be careful to ensure that the model results are consistent with the observed data. For example, say a commuter rail survey yields data on the average length of trips on commuter rail. In this case, for modeled trips that include both commuter rail and bus segments, the length of the commuter rail segment must be considered when comparing to the observed data.

At either level, it is worthwhile for transit trip length comparisons to be segmented using available variables. If the survey data source can provide statistically significant information on trip lengths by trip purpose, traveler/household characteristics (e.g., income level), or subregional geography, it makes sense to perform the comparisons by market segment.

The increasing use of automated passenger counters (APCs) by transit operators might provide an alternate source of data for estimating the in-vehicle portion of observed transit trip lengths. Specifically, if the passengers on-board the transit vehicles between stops are known, the total passenger-miles of travel (PMT) can be estimated. The PMT divided by the total observed boardings provides an estimate of average passenger trip lengths for unlinked trips. This value could be estimated for the region, by service type, by corridor, or by route. Average passenger trip lengths determined using this process can be stratified only by type of service. Stratification by trip purpose or by passenger socioeconomic stratum is not possible.

7.1.3 Disaggregate Checks

As discussed in Section 1.4.3, in disaggregate validation, model predictions are compared with observed data to reveal systematic biases. Disaggregate checks are appropriate for estimated models, as opposed to transferred models where the estimation data set would not be available. Logit models are disaggregately estimated (one record per trip/activity), and therefore disaggregate validation should be performed when logit mode choice models are estimated, along with the aggregate checks described above.

Generally, disaggregate validation is performed by applying the model using a data set with known choice results (such as a revealed-preference survey data set) and checking the results by one or more segmentation variables. Examples of segmentation variables include:

- Income level;
- Vehicle availability level;
- Geographic segmentation (e.g., counties, area types); and
- Trip length segments.

As discussed in Section 6.2.3, disaggregate validation of a model ideally should be performed using a data set that is independent of the data set used for model estimation. However, most urban area household travel surveys have such small sample sizes that the entire data set is needed for model estimation and so there is no independent model estimation data set available for validation. This is especially true for mode choice models, where the household survey itself is often inadequate for model estimation due to low incidence of transit travel.

Limited disaggregate validation can be performed using the same data set used for model estimation, but reporting the results by market segment. Logit model estimation software has the capability to apply the estimated model to a data set in the same form as the estimation data set. For example, a logit mode choice model could be applied to the data set used for estimation but the results may be reported by vehicle availability or income level. It might be found, for example, that transit with auto access is being chosen too often in the model by households with zero vehicles.

Table 7.1 presents an example of disaggregate model validation. This model has five modes, and the results of the application of the estimated model to the estimation data set are reported by vehicle availability level. The "number chosen" represents the choices in the observed data while the "number predicted" represents the model application results. A number predicted that is greater than the number chosen is shown in italics while the asterisks represent the number of standard deviations beyond one that the predicted number differs from the chosen (e.g., one asterisk means that the difference between the number chosen and number predicted is between one and two standard deviations).

Generally, it is desirable to see a good match between the number predicted and the number chosen for all cells. (In the format shown in Table 7.1, this would be represented by a small number of asterisks.) Where there is not a good match, patterns of underestimation or overestimation of demand should be noted. For example, in Table 7.1, the model underestimates nonmotorized demand, but the underestimation is greater for lower vehicle availability levels. One might consider adding a variable representing vehicle availability or revising coefficients of such variables to address this issue. Transit use is overestimated although the sample sizes for the individual vehicle availability levels are very small, making analysis of trends across vehicle availability levels problematic. It is best to perform disaggregate validation multiple times, with different segmentation schemes, to identify the most significant validation issues to address.

Table 7.1 Disaggregate Model Validation Example

Mode Alternative	0 Vehicle	1 Vehicle	2 Vehicles	3+ Vehicles	All
Nonmotorized					
Number chosen	94	210	540	318	**1,162**
Standard deviation	15.0	28.0	53.0	38.0	**72.6**
Number predicted	66	156	530	302	**1,054**
Drive alone					
Number chosen	0	2532	8450	6466	**1,7448**
Standard deviation	0.0	50.8	89.6	71.6	**135.4**
Number predicted	0	*2534*	8408	*6509*	***17,451***
Shared ride					
Number chosen	200	98	372	284	**954**
Standard deviation	26.0	24.0	43.0	35.0	**57.8**
Number predicted	*212*	*136*	*392*	246	***986***
Transit-walk access					
Number chosen	58	56	26	12	**152**
Standard deviation	15.6	14.6	16.2	8.8	**24.2**
Number predicted	*72*	*68*	*54*	*19*	***213***
Transit-auto access					
Number chosen	6	18	18	4	**46**
Standard deviation	4.8	7.2	7.6	4.8	**12.0**
Number predicted	*8*	*20*	*22*	*8*	***58***
Total					
Number chosen	358	2,914	9,406	7,084	**19,862**
Number predicted	358	2,914	9,406	7,084	**19,862**

7.1.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of mode choice. It is desirable to have the aggregate base year mode shares by market segment match the observed data, but it is important to recognize that the "observed" data are created only by combining different data sources such as traffic counts and household and transit on-board survey data, the latter of which represent relatively small samples of the population.

7.1.5 Reasonableness and Sensitivity Testing

Parameter Estimates

Mode choice model parameters, the coefficients and constants in the utility functions, may be estimated using local data, transferred from another model, or asserted. An important reasonableness check is that all mode choice model parameters should be of reasonable sign and magnitude.

Estimated parameters should be checked not only for reasonableness, but also for statistical significance. A complete set of statistical tests should be performed as part of the model estimation process.[23] While this testing is not described in detail here, those checking the reasonableness of estimated parameters should be aware of the statistical significance of the estimates.

It is important to distinguish between generic coefficients, which have the same value for all alternatives, and alternative-specific, or mode-specific, coefficients, which are different for every alternative. Generally, if a variable has the same value for a particular traveler/trip for every alternative (for example, number of autos owned, in a mode choice model), it must be mode-specific (with the coefficient equal to zero for one alternative). It is possible to constrain the coefficients for mode-specific coefficients to be the same for a group of alternatives, but not for all alternatives. A generic coefficient may be used when the value of the variable is not the same for all alternatives (and may be viewed as different variables for each mode). For example, in-vehicle time has different values depending on whether the mode is auto, transit, etc.

For parameters that are mode-specific, the sign should be positive if the variable represents a characteristic that is positively correlated with the use of the mode, and vice versa. For example, if a transit mode utility has a variable that is equal to one if the household owns no vehicles and zero otherwise, the variable should have a positive coefficient (assuming that auto mode utilities do not also have this variable).

It is important to recognize that the values of mode-specific coefficients are relative to those in the utility functions for other modes. The correct sign for a coefficient may be positive if the coefficient of one mode is zero (this is often referred to as the "base" mode) while the correct sign may be negative with a different base mode. An example for a multinomial logit model with three modes is shown in Table 7.2.

---

[23] See, for example, Ben-Akiva, Moshe and Steve Lerman, *Discrete Choice Analysis, Theory and Application to Travel Demand*, The MIT Press, December 1985; or Koppelman, Frank and Chandra Bhat, *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*, prepared for the Federal Transit Administration, 2006.

Table 7.2 Example of Equivalent Logit Model Coefficients

Coefficient Set[a]	Mode A	Mode B	Mode C
1	0	1.5	-1.0
2	-1.5	0	-2.5
3	1.0	2.5	0

[a] If each coefficient set represents the coefficients for the same variable for the three modes in the model, coefficient sets 1, 2, and 3 represent equivalent models.

The determination of "reasonable" requires experience and judgment. One common way of examining reasonableness is to compare the magnitude of model coefficients to those used in other models. However, the values of model parameters depend on model structure, the presence or absence of other variables, and the context of the area being modeled. It is not valid, for example, to assume that the coefficients in a model with three variables would be the same as the coefficients for the same variables in a model with those variables plus three others. It would also be unreasonable to assume that, for example, a cost variable coefficient in a model, which represents the sensitivity of mode choice to, say, one dollar of travel cost, would be the same in another model for an area with a significantly higher cost of living, or even in another model estimated for the same area five or ten years earlier.

Level of service coefficients should always be negative in sign since higher values of the variables (time, cost) for a mode represent a worse level of service. These coefficients represent the sensitivity of mode choice to particular components of level of service. Therefore, they might be expected to have similar values for all mode choice models, at least those structured similarly, since it would seem unlikely that travelers in one urban area are far more or less sensitive to, say, wait time than they are in another area.

Some studies have shown consistency among parameter estimates for various mode choice models in the U.S. Tables 7.3 through 7.5 compare parameters from models estimated from the late 1970s through 1990s.[24] The coefficients for the different models show some consistency, especially for time variables in home-based work models, but also some significant differences. It is unknown how much the differences in coefficients are due to actual differences in travel behavior across different urban areas (and years), and how much are due to model estimation related issues. Many of the models reviewed contained a variety of other variables whose presence in the models could be affecting the estimated

[24] Rossi, T. F., and M. L. Outwater, "How Valid Is It to Transfer Mode Choice Model Parameters?" proceedings of the Seventh National Conference on Transportation Planning Applications, Transportation Research Board, 1999.

values of level of service coefficients. It is important to consider the effects of other variables in the model when evaluating the reasonableness of model coefficients.

Table 7.3 Comparison of Home-Based Work Model Parameters

Model	Year	Auto IVT (Min)	Auto OVT (Min)	Auto Operating Cost ($)	Parking Cost ($)	Transit IVT (Min)	Transit Walk Time (Min)	Transit Wait Time (Min)	Transit Transfer Time (Min)	Transit Cost ($)
Composite[a]		-0.025	-0.050	-0.400	-1.200	-0.025	-0.050	-0.050	-0.050	-0.500
Dallas	1984	-0.030	-0.055	-0.460	-1.160	-0.030	-0.055	-0.055	-0.055	-0.460
Denver	1985	-0.018	-0.093	-0.350	-0.950	-0.018	-0.054	-0.028	-0.059	-0.440
Detroit	1965	-0.046	-0.260	-0.650	-0.650	-0.046	-0.064	-0.117	-0.038	-0.650
Los Angeles	1991	-0.021		-0.296	-0.296	-0.021	-0.053	-0.053	-0.053	-0.296
Milwaukee	1991	-0.016	-0.041	-0.450	-0.450	-0.016	-0.041	-0.041	-0.041	-0.450
Philadelphia	1986	-0.042		-0.260	-0.260	-0.011	-0.032	-0.051	-0.051	-0.115
Pittsburgh	1978	-0.047	-0.069	-2.100	-2.100	-0.047	-0.069	-0.069	-0.069	-2.100
Portland	1985	-0.039	-0.065	-1.353	-1.353	-0.039	-0.065	-0.040	-0.090	-1.353
Sacramento	1991	-0.025	-0.038	-0.279	-0.279	-0.025	-0.038	-0.038	-0.038	-0.279
St. Louis	1965	-0.023	-0.057	-1.170	-1.170	-0.023	-0.057	-0.057	-0.057	-1.170
Tucson	1993	-0.018		-0.184	-0.184	-0.018	-0.040	-0.040	-0.040	-0.184
Average		**-0.029**	**-0.085**	**-0.687**	**-0.805**	**-0.027**	**-0.052**	**-0.053**	**-0.054**	**-0.682**

[a] Schultz, G., Memorandum to Seattle Metro Files, March 5, 1991.

Table 7.4 Comparison of Home-Based Nonwork Model Parameters

Model	Year	Auto IVT (Min)	Auto OVT (Min)	Auto Operating Cost ($)	Parking Cost ($)	Transit IVT (Min)	Transit Walk Time (Min)	Transit Wait Time (Min)	Transit Transfer Time (Min)	Transit Cost ($)
Composite[a]		-0.008	-0.020	-0.800	-2.000	-0.008	-0.020	-0.020	-0.020	-1.000
Dallas	1984	-0.004	-0.007	-0.230	-0.580	-0.004	-0.007	-0.007	-0.007	-0.230
Denver	1985	-0.012	-0.076	-1.310		-0.012	-0.076	-0.076		
Detroit	1965	-0.007		-9.960	-9.960	-0.007	-0.011	-0.018	-0.018	-9.960
Los Angeles	1991	-0.024	-0.061	-0.216	-0.216	-0.024	-0.061	-0.061	-0.061	-0.216
Milwaukee	1991	-0.009	-0.069	-1.330	-1.330	-0.009	-0.069	-0.069	-0.069	-1.330
Philadelphia	1986	-0.020		-0.100	-0.100	-0.001	-0.002	-0.002	-0.002	-0.012
Pittsburgh	1978	-0.017	-0.079	-1.450	-1.450	-0.017	-0.079	-0.079	-0.079	-1.450
Portland	1985	-0.033	-0.086	-0.399	-0.399	-0.033	-0.086	-0.086	-0.086	-0.399
Sacramento	1991	-0.021	-0.055	-0.557	-0.557	-0.021	-0.055	-0.055	-0.055	-0.557
St. Louis	1965	-0.024	-0.060	-2.430	-2.430	-0.024	-0.060	-0.060	-0.060	-2.430
Tucson	1993	-0.024		-0.250	-0.250	-0.024	-0.054	-0.054	-0.054	-0.250
Average		**-0.020**	**-0.068**	**-1.855**	**-1.855**	**-0.018**	**-0.053**	**-0.054**	**-0.054**	**-1.845**

[a] Schultz, G., Memorandum to Seattle Metro Files, March 5, 1991.

Table 7.5 Comparison of Nonhome-Based Model Parameters

Model	Year	Auto IVT (Min)	Auto OVT (Min)	Auto Operating Cost ($)	Parking Cost ($)	Transit IVT (Min)	Transit Walk Time (Min)	Transit Wait Time (Min)	Transit Transfer Time (Min)	Transit Cost ($)
Composite[a]		-0.020	-0.050	-0.600	-1.600	-0.020	-0.050	-0.050	-0.050	-0.800
Dallas	1984	-0.012	-0.024	-0.440	-0.700	-0.012	-0.024	-0.024	-0.024	-0.440
Denver	1985	-0.013	-0.033	-1.330		-0.013	-0.033	-0.033		
Detroit	1965	-0.016	-0.355	-4.670	-4.670	-0.016	-0.023	-0.039	-0.039	-4.670
Los Angeles	1991	-0.050	-0.126	-0.453	-0.453	-0.050	-0.126	-0.126	-0.126	-0.453
Milwaukee	1991	-0.011	-0.074	-0.310	-0.310	-0.011	-0.074	-0.074	-0.074	-0.310
Philadelphia	1986	-0.004	-0.009	-0.046	-0.114	-0.007	-0.017	-0.017	-0.017	-0.086
Pittsburgh	1978	-0.012	-0.195	-3.050	-3.050	-0.012	-0.195	-0.195	-0.195	-3.050
Portland	1985		-0.127				-0.127	-0.127	-0.127	
Sacramento	1991	-0.035	-0.082	-1.103	-1.103	-0.035	-0.082	-0.082	-0.082	-1.103
St. Louis	1965	-0.023	-0.058	-2.350	-2.350	-0.023	-0.058	-0.058	-0.058	-2.350
Tucson	1993	-0.014		-0.151	-0.151	-0.014	-0.031	-0.031	-0.031	-0.151
Average		**-0.020**	**-0.128**	**-1.517**	**-1.525**	**-0.021**	**-0.081**	**-0.083**	**-0.083**	**-1.522**

[a] Schultz, G., Memorandum to Seattle Metro Files, March 5, 1991.

It is important to consider the coefficients not only individually, but also the relationships between them. In nearly all mode choice models, coefficients for variables representing out-of-vehicle time – including wait, walk access/egress, and transfer time – are greater in absolute value than in-vehicle time coefficients. This relationship implies that time spent waiting or walking is considered more onerous than time spent in a vehicle, usually sitting. As Tables 7.2 through 7.4 show, the ratios of out-of-vehicle time coefficients to in-vehicle time coefficients are about 2 to 3 for home-based work trips with some higher values estimated for nonwork trips. (Note that in some models, different components of out-of-vehicle time have different coefficients.)

Another relationship that can be checked is the value of (in-vehicle) time, which is represented by the ratio of the in-vehicle time coefficient to the cost coefficient. Represented in dollars per hour, the values of time range from about $2 to $5 per hour for work trips, and 50 cents to $5 per hour for nonwork trips with a few outliers for each trip purpose and greater variation for nonwork trips. (It should be noted that cost coefficients and values of time represent the specific base years for which the models were estimated, and the tables have not normalized the cost coefficients to represent the same year's dollar value.)

The row labeled "Composite" in Tables 7.2 through 7.4 shows coefficient values derived by Schultz in 1991[25] as cited by Cambridge Systematics, Inc. and Barton-Aschman Associates, Inc.[26] Schultz used coefficients from mode choice models estimated in the 1970s to develop composite coefficients for use in other areas. As Tables 7.2 through 7.4 show, most of the "composite" coefficients are within the range of experience from the more recent models cited in the tables, and many of them are close to the averages of the coefficients shown in the tables.

Schultz's work helped inform the development of guidelines by the Federal Transit Administration (FTA) for ridership forecasts for Section 5309 New Starts projects. At a workshop in June 2006,[27] FTA presented guidelines for mode choice coefficient values, including the following:

- A typical range for the value of the in-vehicle time coefficient for home-based work trips is -0.03 and -0.02. If the coefficient falls outside the range, FTA says that "some further analysis (is) appropriate."

- In-vehicle time coefficient for nonhome-based trips should approximately be the same as the in-vehicle time coefficient for home-based work trips.

- A typical range for the in-vehicle time coefficient for home-based other nonwork trips is 0.1 to 0.5 times the in-vehicle time coefficient for home-based work trips.

- A typical range for the coefficient of out-of-vehicle time is 2 to 3 times the corresponding coefficient for in-vehicle time. FTA believes that "compelling evidence" is needed to justify ratios outside this range.

If a nested logit mode choice formulation is used, a logsum variable is included in the model specification for each nest of modal alternatives. The coefficients of these variables are estimated or asserted. While there are no specific reasonableness checks of logsum variable coefficients, especially asserted coefficients, the coefficients' validity must be checked with respect to two rules:

- Logsum coefficients must be between zero and one. The coefficients should be statistically different from both zero and one (although statistical significance can be checked only for estimated coefficients, not for asserted coefficients).
- The logsum coefficient for a nest should be lower than the logsum coefficient for any higher level nest of which the nest is a component.

---

[25] Schultz, G., Memorandum to Seattle Metro Files, March 5, 1991.

[26] Cambridge Systematics, Inc., and Barton-Aschman Associates, Inc., *Short-Term Travel Model Improvements*, prepared for the U.S. Department of Transportation Travel Model Improvement Program, October 1994.

[27] Federal Transit Administration, "Travel Forecasting for New Starts Proposals," Minneapolis, Minnesota, June 15-16, 2006.

Mode-specific constants are also model parameters that should be checked for reasonableness. Checks of constants are discussed in Section 7.1.6.

Sensitivity Testing

Sensitivity testing can be performed for mode choice models by varying model inputs and checking results for reasonableness. Model inputs that can be varied include level of service variables used in the trip distribution model (time/speed and cost) and the demographic- or zone-level variables that are used as model inputs. Some example tests include:

- Increasing or decreasing highway or transit travel times by a fixed percentage regionwide;
- Increasing/decreasing parking costs in the CBD by a fixed percentage;
- Increasing/decreasing headways on selected transit routes or submodes by a fixed percentage or amount;
- Increasing/decreasing fares on selected transit submodes by a fixed percentage;
- Changing development patterns for forecast years by moving projected new activity among different parts of the modeled region (e.g., from suburbs to small urban centers or from outlying areas to infill); and
- Reallocating the number of households by income level for a forecast year.

The resultant changes in demand due to changes in a model input variable reflect the sensitivity to the variable; the sensitivity level is determined by the coefficient of the variable in the utility function. Simple "parametric" sensitivity tests can be performed by introducing small changes in the input variable or in the parameter itself and checking the results for reasonableness.

The changes in demand for a modal alternative (or group of alternatives) with respect to a change in a particular variable can be expressed as arc elasticities. Arc elasticity may be calculated as:

$$\eta = \frac{\ln(q_2) - \ln(q_1)}{\ln(p_2) - \ln(p_1)} \qquad (7.2)$$

Where:

η = Arc elasticity;

q_1 = Value of result (demand) for base condition;

q_2 = Value of result (demand) for change condition;

p_1 = Value of variable for base condition; and

p_2 = Value of variable for change condition.

While there are some rules of thumb for what constitute reasonable elasticities,[28] there are not yet specifically defined ranges of reasonable elasticities. Generally, experience has shown that elasticities of transit demand with respect to level of service variables are usually well under 1.0 in absolute value. The "Simpson-Curtin Rule" indicates that the elasticity of transit demand with respect to fare is about -0.3. It is important to recognize that since the logit formulation is nonlinear, the elasticities of modal demand are not constant. The elasticity calculated for one particular "point" (say, a specific market segment defined geographically, demographically, and temporally) will not be equal to the elasticities computed at other points.

7.1.6 Troubleshooting Strategies

Issues discovered during the model checks described in Sections 7.1.2, 7.1.3, and 7.1.5 may imply errors in mode choice model parameters, input data (networks/skims or trip tables), or highway or transit path building procedures. Table 7.6 shows some of the typical problems that may be evident from these tests. Note that while four-step model terminology (e.g., "trip distribution") is used in the table, if a mode choice model that is part of an activity-based model system is being validated, the corresponding components (e.g., tour- and trip-level destination choice) should be used.

The types of actions shown in Table 7.6 show some similar themes for each mode. If the modeled base year demand for a mode within a market segment (defined by a combination of geographic, temporal, and/or demographic segmentation) is too high or too low compared to the observed demand estimate, it is important to make sure that the total demand for the segment is correct. For example, if a corridor has too much modeled transit demand for the a.m. peak period, the total demand in the corridor should be checked. Say the observed transit demand estimate is 1000 and the total modeled demand for the segment is 2000. The analyst should determine whether it is reasonable for the transit share to be 50 percent for the segment. If not, and the observed transit demand estimate is accurate, then it is likely that the total modeled demand from the trip distribution process is inaccurate. In this case, calibration procedures focused on mode choice will not only be ineffective, but they may also possibly be counterproductive, introducing counterbalancing errors to offset the trip distribution error. While this may "fix" the base year results, sensitivities for future year forecasts may be adversely impacted.

---

[28] See, for example, TCRP Report 95, Traveler Response to Transportation System Changes, Transportation Research Board, Washington, D.C.

Table 7.6 Troubleshooting Strategies for Issues with Mode Choice Model Results

Issue	Potential Troubleshooting Strategies
1. Transit demand for specific market segments is too high or low	• Check trip distribution to determine if overall travel in the market is correct • Check implied transit share for market (observed transit demand divided by the total travel in market from trip distribution) to determine if it is reasonable • Recheck transit skim data related to the market (for appropriate geographic scope and time periods) • Consider revisions to logit model structure • Consider adding or removing indicator variables related to the market, or adjusting the coefficients of existing indicator variables (see discussion below) • Consider revisions to alternative specific constants (see discussion below)
2. Nonmotorized mode shares for specific market segments are too high or low	• Check trip distribution to determine if overall travel in the market is correct • Recheck skim data (usually distance skims) related to the market (for appropriate geographic scope and time periods) • Consider adding or removing indicator variables related to the market, or adjusting the coefficients of existing indicator variables (see discussion below) • Consider revisions to alternative specific constants (see discussion below)
3. Auto submode (e.g., toll versus nontoll) shares for specific market segments are too high or low (note that vehicle occupancy checks are discussed separately in Section 7.2)	• Check trip distribution to determine if overall travel in the market is correct • Check implied mode share for market (observed modal demand divided by the total travel in market from trip distribution) to determine if it is reasonable • Recheck skim data related to the market (e.g., toll cost skims for toll road corridors, for appropriate geographic scope and time periods) • Consider adding or removing indicator variables related to the market, or adjusting the coefficients of existing indicator variables (see discussion below) • Consider revisions to alternative specific constants (see discussion below)
4. Mode choice too sensitive or insensitive to level of service changes	• Adjust parameters for appropriate level of service variables in utility functions

If the validation issues are found to be most likely associated with the mode choice model, the best calibration method may not immediately be apparent from among several possibilities. The best strategy for addressing the issue might lie in:

- Revisions to input data or the procedures to create them (e.g., skimming);
- Changes to model parameters (utility equation coefficients or alternative specific constants);
- Changes to the model structure (e.g., nesting);
- Changes to modal alternative definition; or
- Changes to market segmentation in the model.

The tests described in Table 7.6 are not intended to be performed in isolation, and an iterative approach to testing may be necessary. For example, consider a model where both modeled transit and modeled nonmotorized mode shares are too low. Adjustments that increase the transit shares may further decrease the nonmotorized mode shares; when calibration actions are taken to increase the nonmotorized mode shares, the transit shares may be affected further. If it is found that overall trip distribution issues are responsible for mode choice validation issues, they will affect all modes.

Mode-Specific Constants

The interpretation of a mode-specific constant is that it represents the part of the modal utility that is not considered by the variables in the utility function. The variables represent measurable characteristics of the trip/tour, the traveler, and the area on which the trip/tour is made that affect the choice of mode. The constant, therefore, represents the sum of items that affect the choice that are not included in the variables. These items may include reliability, comfort, convenience, safety, and many other factors.

In model estimation, the original values of constants are estimated. The constants can easily be revised so that modeled mode shares match targets. It is evident that the "correct" values for modal constants are unknown since they represent factors affecting choice that could not be quantified sufficiently to be included in other mode variables. It would be incorrect, however, to assume that all validation issues are the result of these unknown factors. As is the case with K-factors, as discussed in Section 6.2.6, simple adjustments to modal constants estimated using weighted samples should be considered "a last resort" after all other possible causes for error and calibration adjustments have been considered, and so this is why they are listed as the last items in each box in Table 7.6. Because constants can be revised to provide nearly perfect matches between modeled and observed mode shares, it can be very tempting to revise modal constants to resolve differences in shares without determining whether it is the best method to solve the problem at hand.

It should be noted that mode choice constants that are estimated from nonweighted data sets do not reflect the maximum likelihood estimates of the observed behavior and therefore should be adjusted after model estimation. Similarly, if constants are transferred from another model, they will need to be adjusted since they do not represent the observed behavior in the application context. The "last resort" nature of adjusted constants applies only to models estimated using weighted data sets.

The values of mode-specific constants, whether estimated or revised during calibration, should be checked for reasonableness. One way of doing this is to compare the value of a constant relative to the constants of other modal alternatives to the values of other parameters. For example, the difference between the rail and bus constants could be divided by the in-vehicle time coefficient to express the different in units of minutes of in-vehicle time. If the difference between two constants was -0.5 (with the rail constant higher), and the in-vehicle time coefficient was the same for the two modes and equal to -0.025, the difference in the constants is equivalent to -0.5/-0.025 = 20 minutes of in-vehicle time. This implies that all other things being equal, a traveler would be indifferent between a bus trip and a rail trip that is 20 minutes longer.

The interpretation of differences between constants can be muddied somewhat by modal availability issues. For example, it is common to see transit constants that are so much lower than auto constants that it is implied that a traveler would be indifferent between a transit trip and an auto trip that is several hours longer. However, many travelers may not have the auto mode available while others do not consider transit as a viable mode.

Market Segment Indicator Variables

Indicator variables, or dummy variables, take values of zero or one depending on whether the trip/tour or traveler belongs to a particular market segment. Indicator variables may include variables that represent:

- Whether the traveler belongs to a particular demographic group (income level, number of vehicles available to household, etc.);
- Whether the trip or tour is to, from, or through a particular type of area, such as the central business district (CBD);
- Whether the trip or tour has a particular characteristic related to the transportation service provided, such as transit transfers or transfer mode; or
- Whether the trip or tour is for a particular subpurpose.

Since the value of the indicator variable is the same for each market segment represented, it effectively changes the constant term for that market segment by the amount of the indicator variable's coefficient. For example, if a mode has a constant of -0.5, and the traveler is a member of a segment with an indicator variable coefficient of 1.0, the effective modal constant for the traveler is +0.5. The set of effective mode-specific constants for a traveler who is a member of one segment might be quite different than the set for a traveler in another segment.

One check of the coefficients of indicator variables that are ordered, such as income or vehicle availability levels, is that they have a logical progression in the utility of a mode. For example, if there were variables representing households with 0 vehicles, 1 vehicle, 2 vehicles, and 3 or more vehicles, one would usually expect that the coefficient of the 0-vehicle variable in the utility of a mode such as bus with walk access would be the highest of the four, with the coefficients decreasing as the vehicle availability level increases.

There might be some modes for which the highest coefficient might not be representing one of the extreme category values. For example, a "bus with auto access mode" might have a higher coefficient for 1-vehicle households than for 0, 2, or 3+ vehicle households. In such a case, however, the coefficients should be consistently decreasing from the highest value, or increasing from the lowest, on both sides of the maximum (or minimum). If the 1-vehicle coefficient is highest, then the 3+ vehicle coefficient should be lower than the 2-vehicle coefficient.

It is important to recognize that indicator variables, especially those that represent categories of continuous or near-continuous variables, such as area types or income levels, may be susceptible to the "cliff" problem described in Section 3.3. A small change in income or a change in trip origin/destination across an area type boundary could result in a large change in utility for a mode. It is better to use continuous variables where possible although it is not always feasible – for example, the survey data used for model estimation or validation may have reported income only for a small number of categories.

It is just as easy to revise coefficients of indicator variables as it is to revise mode-specific constants, and it is just as tempting to make such revisions to attempt to match aggregate observed mode shares for specific market segments by doing so. It is therefore also considered a "last resort" to make such changes in model calibration.

7.1.7 Forecasting Checks

Mode choice forecasts should be compared to base year estimates for reasonableness. Specifically, mode shares by market segment as well as overall regional shares for the forecast year should be compared to the base year. As is the case with the aggregate checks described in Section 7.1.2, market segments should include trip or tour purposes as well as demographic segments, such as income or vehicle availability levels, and geographically defined segments.

It is not necessarily the case that forecasted mode shares should be similar to modeled base year shares. The most obvious reason for differences might be changes to the transportation system in the forecast year, such as new or extended (in space or time) transit service that would increase transit mode shares. However, it is equally important to consider the effects of changes in land use patterns or demographic changes for residents of the modeled region. For example, if new development is concentrated in outlying areas not served by transit, regional mode shares may decrease while mode shares in the part of the region served by transit remain stable. If income or vehicle availability levels are projected to increase in the future, transit shares may decline.

By performing the aggregate comparisons for a variety of segmentation schemes, it should be possible to obtain a "picture" of the reasons for changes in forecasted mode shares. The examples above demonstrate the value of performing these checks for a comprehensive set of segmentation schemes.

Some checks may reveal an issue with the mode choice model that might require additional testing with the base year scenario. For example, say that the transit share declines substantially while income and vehicle availability increase slightly. This may indicate that the model is too sensitive to one or both of these variables. Additional sensitivity checks using the base year scenario, where all other variables can be kept constant, could be valuable in identifying the problem and suggesting corrective actions.

As discussed above, the mode choice model validation process is tied in with the transit assignment validation process. It is worthwhile to consider performing forecasting checks of both mode choice and transit assignment before determining the best actions to take to address any issues identified.

7.2 Checks of Vehicle Occupancy Results

Vehicle occupancy levels can be either results of mode choice models where occupancy levels are alternatives (single occupant, two occupants, etc.) or exogenously derived factors applied to auto person trips or tours. In either case, the numbers should be checked for reasonableness and compared to observed data. Usually, vehicle occupancy is considered separately by trip or tour purpose.

It should be noted that even for mode choice models with alternatives corresponding to specific occupancy levels, an exogenously derived factor must be applied to the auto person trips corresponding to the highest occupancy level modeled. For example, if the highest level modeled is "three or more persons," an exogenously derived factor must be used to convert these trips (or tours) to auto vehicle trips (tours). This factor must be greater than or equal to three and reflects an assumed number of vehicles with three persons, four persons, five persons, etc.

Therefore, in mode choice models with separate vehicle occupancy-level alternatives, the shares for each occupancy level should be validated in the same way that other mode choice results are validated, as described in Section 7.1, and the exogenously derived factors for the highest levels should be checked as well. For models in which only exogenous derived (i.e., nonmodeled) vehicle occupancy factors are used, the factors derived from other sources are checked during validation.

In activity-based models, the specific definitions of tour-level modes are important in determining vehicle occupancy levels. Most tour-level mode choice models include separate alternatives at least for single occupant and multiple occupant vehicle tours, with some models including separate alternatives for higher numbers of occupants. Most commonly, a tour is defined as single occupant only if there is only one occupant (the driver) for the entire tour. If a passenger is picked up or dropped off during the tour, the tour is defined as multiple occupant. Most activity-based models have an "escort" tour purpose; the single occupant mode is by definition unavailable for escort tours.

7.2.1 Sources of Data

The main sources of data for validation of vehicle occupancy levels include the following:

- **Household travel/activity survey –** If such a survey is available, it may have also been a data source for model estimation. The expanded household survey data can be used to produce observed vehicle occupancy levels by purpose for a number of geographic and demographic market segments.

- **Census data –** The CTPP, now based on the ACS, contains information on modes for work travel. As discussed in Section 7.1.1, these data are more consistent with tour-level work travel than trip level, and data on travel only in the home to work direction are available.

- **National sources –** Perhaps the most comprehensive source for vehicle occupancy data is NHTS. The 2010 update to NCHRP Report 365, Travel Estimation Techniques for Urban Planning, will also provide information on vehicle occupancy levels for specific urban area types.

- **Field observations –** While observed vehicle occupancy data are usually available only for a few locations, they may be useful in some cases. One example is observed data at external stations, which may be used in models where external person trips are modeled.

7.2.2 Aggregate Checks

The aggregate checks for vehicle occupancy are checks of the shares of person trips or tours falling into each occupancy category, if the categories are modeled as alternatives in mode choice, and checks of the overall vehicle occupancy levels, including the levels for the highest categories appearing as alternatives in the mode choice model.

It is important to understand the correct way to compute vehicle occupancy and to do so consistently for the model results and the observed data to which they are compared. The vehicle occupancy level is obvious at the disaggregate level; it is simply the number of persons in each vehicle. The average vehicle occupancy for a group of travelers, however – say all home-based work trips – may be computed in more than one way. Consider an example where there are six vehicles: three with one person, two with two persons, and one with three persons. This means that three persons are in single occupant vehicles, four are in two occupant vehicles, and three persons are in three occupant vehicles. The shares represented in a mode choice model of person trips would be 30 percent single occupant, 40 percent two occupant, and 30 percent three occupant. This implies an average vehicle occupancy of 2.0, which is clearly incorrect: there are 10 persons in six vehicles, and the actual vehicle occupancy is 1.67.

In other words, the incorrect average vehicle occupancy is obtained when one multiplies the number of person trips for each occupancy level by the occupancy level and divides the sum of these trips by the total number of person trips. The correct method is to estimate the implied vehicle trips by dividing the number of person trips for each occupancy level by the occupancy level and sum the result over all occupancy levels. The average vehicle occupancy level is then computed by dividing the total person trips by the implied vehicle trips. For the example above, the implied number of vehicle trips is given by:

(3 pers trips / 1 pers/veh) + (4 pers trips / 2 pers/veh) + (3 per trips / 3 pers/veh)

which yields 6 implied vehicle trips. The average vehicle occupancy is 10 person trips divided by 6 implied vehicle trips, or 1.67.

In person-based or person trip-based models, person trips or tours are generated and eventually converted to vehicle trips. If a vehicle trip-based model is being used, there is no conversion required; if one wanted to compute the number of person trips from the number of person trips, the vehicle-based occupancy level would be used.

Occupancy-Level Shares

The most basic check is to compare the modeled base year model shares of trips made by vehicle occupancy, both by trip or tour purpose and for all trips/tours, to observed shares. In most cases the best (or only) source of observed occupancy levels is from the household survey data. The modeled occupancy levels should match the observed data closely as the percentage error in vehicle occupancy results in the same level of error in the number of vehicle trips. For example, if the overall vehicle occupancy level is too low by five percent, the number of vehicle trips is too high by five percent. (It should be noted, of course, that there is error inherent in the observed data, which is based on a sample, even though it is impossible to precisely know the level of error in the survey data set.)

When a sufficient household survey data set is not available, modeled occupancy levels may be compared to representative data from another data set, such as the NHTS. In these cases, of course, the survey data do not truly represent observed data for the modeled area, and so a precise match is not necessary. The comparison represents more of a reasonableness check.

Average Vehicle Occupancy

Traditionally, average vehicle occupancies have been reported by trip purpose. However, it is important to recognize that the trip or activity purposes of all passengers in a vehicle are not necessarily the same. For example, in one region, the home-based work average vehicle occupancy has been reported as 1.14 persons per vehicle. Such a report is not strictly correct since the trip purposes of the other passengers in the vehicles may not be home-based work. The correct report is that home-based work person trips are made in vehicles with an average of 1.14 persons per vehicle.

Overall vehicle occupancy should be checked for all person-based models, by trip/tour purpose and for all trips/tours. The average vehicle occupancy should be computed by dividing the total person trips for each market segment by the implied number of vehicles used for those trips.

7.2.3 Disaggregate Checks

If a mode choice model with alternatives corresponding to specific occupancy levels has been estimated, disaggregate validation of the logit model should be performed as part of the validation of the estimated model, as described in Section 7.1.3.

7.2.4 Criteria Guidelines

There are no applicable criteria guidelines for checks of vehicle occupancy.

7.2.5 Reasonableness and Sensitivity Testing

The reasonableness checks for vehicle occupancy are the comparisons to traffic or person trip counts, as discussed in Section 7.2.2.

Sensitivity testing of vehicle occupancy can be performed for a mode choice model with alternatives corresponding to specific occupancy levels. The types of tests discussed in Section 7.1.5 can be performed, using available model variables that may affect vehicle occupancy, such as household size and HOV lane travel times.

7.2.6 Troubleshooting Strategies

Table 7.7 shows some typical issues with vehicle occupancy results and suggested strategies for dealing with them.

Table 7.7 Troubleshooting Strategies for Issues with Vehicle Occupancy Results

Issue	Potential Troubleshooting Strategies
1. Modeled vehicle occupancy by trip/ tour purpose differs significantly from observed levels.	• Check observed data for errors • If a mode choice model with alternatives corresponding to specific occupancy levels is being used, check sensitivity to mode choice model input variables and consider adjusting logit model parameters • Consider adding or removing indicator variables related to the market, or adjusting the coefficients of existing indicator variables (see discussion in Section 7.1.6) • Consider revisions to alternative specific constants (see discussion in Section 7.1.6)
2. Overall modeled vehicle occupancy differs significantly from observed levels.	• Compare modeled shares of trips/tours by purpose to observed

7.2.7 Forecasting Checks

Forecasted vehicle occupancy levels and averages should be compared to base year values for reasonableness. In mode choice models with alternatives corresponding to specific occupancy levels, the forecasted vehicle occupancy levels may differ slightly from the base year values due to changes in the values of forecast year input variables, including household size and levels of congestion on general purpose roadways when HOV facilities are present. In models where fixed vehicle occupancy factors are used, there will be no change in vehicle occupancy levels for the segments (such as trip purposes) over which the actors are applied although there might be small changes in overall vehicle occupancy if the relative numbers of trips in each segment change over time.

8.0 Time of Day

8.0 Time of Day

Many travel demand models have a time-of-day component. It is typical for models to start by estimating daily travel. In a four-step model, the trip generation model is typically applied to estimate average weekday trips. In an activity-based model, one of the first steps is typically a daily activity pattern model, also for weekdays.

It is desirable for many reasons to estimate travel by time of day, including the need for temporally varying model outputs (for example, speeds by time of day for air quality conformity analysis) and to enhance model accuracy (shortest travel paths or transit service may vary between peak and off-peak periods). To do this, daily travel measures are converted to measures by time of day at some point in the modeling process. In most models, a discrete number of time periods are used. Typically, a four-step model with time-of-day modeling uses three to five periods (for example, a.m. peak, mid-day, p.m. peak, evening) while an activity-based model uses more periods, often an hour or half hour (or shorter) in length.

It is important to consider how to determine the period in which a trip or half tour occurs, especially if it begins in one period and ends in another. The three basic ways to assign a trip to a time period are:

1. Based on the departure time;
2. Based on the arrival time; and
3. Based on the temporal midpoint of the trip.

The specific definition usually makes little difference in the percentages of trips occurring in each period, but the definition must be known in order to estimate and validate the model.

Fixed Factor Methods

The most common method of time-of-day modeling in four-step models is simple factoring. At some point in the modeling process, fixed factors specific to trip purpose and direction are applied to daily trips to obtain trips for each time period. While this method is relatively easy to implement and to apply, it is not sensitive to varying transportation levels of service, limiting its usefulness in analyzing policy changes or congestion management activities.

The ways in which fixed time-of-day factors may be applied within the four-step process[29] are:

- In **pre-distribution** application, the daily trips are factored between the trip generation and trip distribution steps of the model process. The data required include factors representing the percentage of trips by purpose during each hour and for each direction, production-to-attraction or attraction-to-production as well as directional split factors. It should be noted, however, that the directional split factors cannot be applied until after both ends of trips have been determined, i.e., after trip distribution. An advantage of this method is that differences in travel characteristics by time of day can be considered in both trip distribution and mode choice. In models with feedback loops, this method can provide a "clean" way to feed back travel times from one iteration to the next; trip distribution, mode choice, and trip assignment can be run separately for each time period, since the factors are applied prior to these steps.

- In **post-distribution** application the factors are applied between the trip distribution and mode choice steps. The data required for this approach to splitting includes factors representing the percentage of trips by purpose during each period and for each direction, production-to-attraction or attraction-to-production. This process also provides an opportunity to consider that some trips are in the attraction-to-production direction and to use skims that reflect correct directionality. However, the modeler should decide whether the additional complexity introduced by doing so is worthwhile.

- In **post-mode choice** application, the factors are applied to daily trips between mode choice and the assignment steps. The data required include factors representing the percentage of the trips by purpose and mode during each time period and for each direction, production-to-attraction or attraction-to-production. An issue with this approach is that the transit path building procedures may not be consistent between mode choice and transit assignment, since mode choice would be done on a daily basis while transit assignment would be done by time period.

- In **post-assignment** application, the factors are applied to loaded trips after the assignment step is complete. The data required include factors that represent the percentage of daily traffic or transit ridership for each time period on a link and can also include directional split factors depending on how the link-level factor is represented. The main limitation of this type of procedure is that equilibrium highway assignment on a daily basis is much less meaningful than assignment for shorter, more homogeneous periods. Also, changes in land use that could affect temporal distribution of traffic are not considered when using fixed link-based factors.

[29] Time-of-Day Modeling Procedures State-of-the-Art, State-of-the-Practice, TMIP, U.S. Department of Transportation, October 1997, Page ES-3.

Time-of-Day Choice Models

As mentioned above, a shortcoming of fixed factoring methods is that time-of-day choice is insensitive to transportation level of service. There has been research into time-of-day choice models that include variables representing level of service, but their success when used with four-step models has been limited. It is difficult to model time-of-day choice where individual trips are modeled independently of all other trips (for example, when the timing of the trip from work cannot be linked to the timing of the trip to work).

Tour- and activity-based models provide a much better way to understand temporal choices and the ways in which relationships between trips can affect these choices. In modern activity-based models, time of day is modeled at a more disaggregate level, for periods of an hour or half hour (or even shorter), than for typical fixed factor models although these shorter periods are often aggregated to a smaller number of longer periods for trip assignment. Time-of-day choice models can take advantage of a variety of person, household, and trip characteristics that affect time-of-day choice as well as information about entire tours or activity patterns. For example, the time of travel to work is surely related to the time of travel home from work, and a tour-based model can use information on both directions of the tour in estimating time of day.

In these advanced models, time of day is estimated at the tour level, usually using a multinomial logit model. Each alternative is a combination of the beginning and ending times, often defined as the start and end times for the primary activity of the tour. This would imply that the combination of ARRIVAL time for the trip TO the primary activity and the DEPARTURE time for the trip FROM the primary activity is being modeled. This type of model can have hundreds of alternatives. In many models, the times of intermediate stops to and from the primary activity are also estimated, again using multinomial logit formulations.

Period to Hour Conversion Factors

It is customary to express capacity as a highway network attribute in terms of vehicles per hour. In models where daily (weekday) travel is modeled without time-of-day considerations be modeling explicitly, it is necessary to attain consistency between the daily vehicle trip tables and the hourly capacity estimates when performing traffic assignment, since volume-capacity relationships are generally used to estimate travel times under congested conditions. This is most commonly done using factors that can be applied to convert the hourly capacity to effective daily capacity (or, conversely, to convert daily trips to hourly trips, which is equivalent mathematically). These factors consider that travel is not uniformly distributed throughout the day and that overnight travel demand is low. The conversion factors are therefore often in the range of 8 to 12, as opposed to 24, which would be the theoretical maximum for an hourly to daily factor.

These conversion factors continue to be needed in models where time periods greater than one hour in length are used, but they convert the hourly capacity to the capacity for the appropriate time period. For example, if an a.m. peak period is defined as 6:00 a.m. to

9:00 a.m., the conversion factor will convert hourly capacity to capacity for the three hour period. It is still important to consider that travel is not uniformly distributed throughout the three hour period although it is likely to be more evenly distributed over a shorter time period, especially a peak period which is likely to be relatively congested throughout. The theoretical maximum for the factor is the number of hours in the period (three, in this example), and in a period where there is roughly uniform congestion throughout the peak period, the factor could be close to 3. Typical factors would range from 2 to 3. The factors for longer off-peak periods would likely be well lower than the theoretical maximum.

8.1 Sources of Data

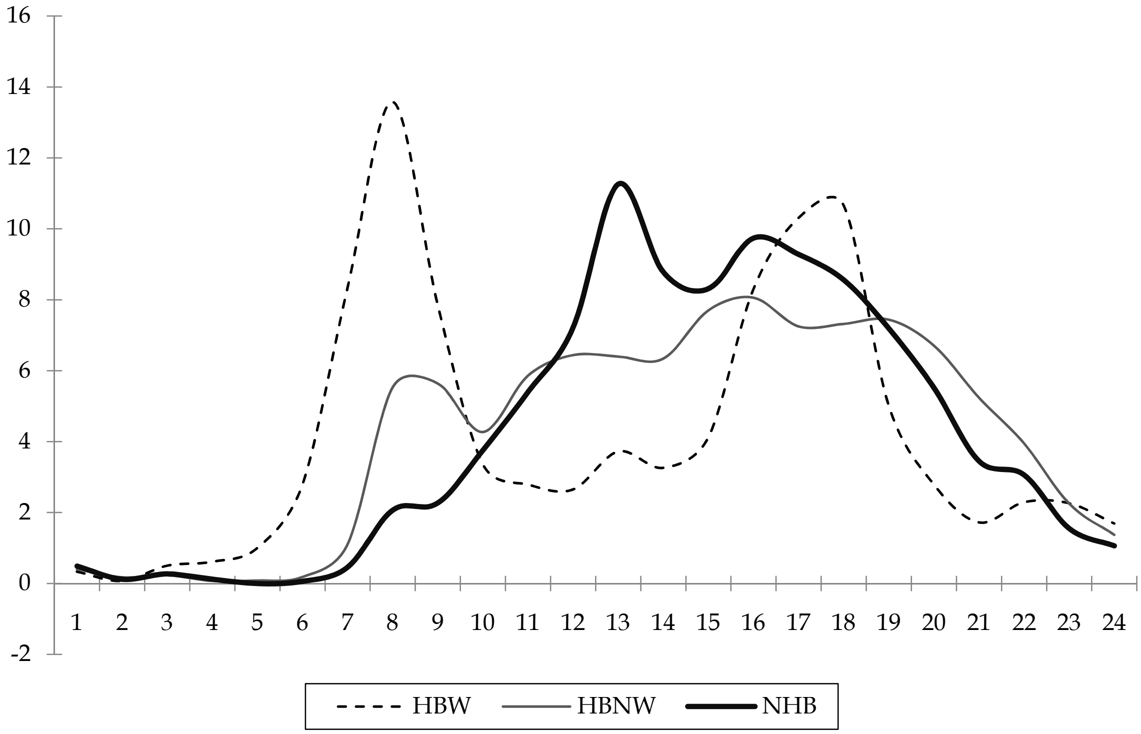
The basic data required for estimating time-of-day models of any type are household survey data, specifically the reported beginning and ending times of activities, tours, and trips.[30] The survey data are processed for the specific type of model being estimated (fixed factor, logit, etc.) and are used separately by trip/tour purpose. These survey data (in expanded form) are also valuable for time-of-day model validation although, as is the case anytime when the estimation data set is used for validation, the data must be used with caution.

For areas without local household survey data, factors from other sources, such as the National Household Travel Survey (NHTS) or NCHRP Report 365 or its update may be transferred. However, as discussed below, time-of-day distributions vary significantly by area, and so significant checking is required when using transferred time-of-day data.

Diurnal travel distributions are typically represented as percentages (for hours or shorter periods) of the total daily travel by purpose, perhaps by mode. Figure 8.1 provides an example of diurnal distributions for three purposes: home-based work (HBW), home-based nonwork (HBNW), and nonhome-based (NHB). However, while distributions such as those shown in Figure 8.1 are typical, they may vary significantly among urban areas that might be considered similar. Table 8.1 shows the percentage of daily travel by purpose occurring during two periods – 7:00 to 9:00 a.m. and 3:00 to 6:00 p.m. – for nine urban areas with populations of approximately 1 million according to the 2000 U.S. Census. While the averages presented in this table, based on data from the 2001 NHTS, have associated statistical error ranges not presented here, it is clear that the percentages for some areas differ significantly from those for other areas. For example, the reported percentage of daily home-based work travel between 3:00 and 6:00 p.m. was nearly twice as high in Providence as in Memphis.

[30] An exception is for post-assignment techniques, where diurnal distributions of traffic volumes and/or transit riders by time of day are used.

Figure 8.1 Example Diurnal Distribution of Trips by Trip Purpose for Large Urban Areas



Source: NCHRP Report 365.

Table 8.1 Time-of-Day Percentages for Urban Areas of Approximately 1 Million in Population

	HBW		HBNW		NHB		All Trips	
	7-9 a.m.	**3-6 p.m.**	**7-9 a.m.**	**3-6 p.m.**	**7-9 a.m.**	**3-6 p.m.**	**7-9 a.m.**	**3-6 p.m.**
Austin	32.3%	20.8%	12.5%	23.8%	6.9%	24.6%	13.6%	23.7%
Buffalo	23.7%	26.7%	9.3%	23.6%	5.9%	23.6%	9.7%	23.8%
Greensboro	30.3%	24.0%	12.2%	25.6%	8.1%	26.7%	12.7%	25.8%
Jacksonville	29.6%	24.7%	10.4%	24.4%	9.1%	27.1%	11.6%	25.3%
Hartford	26.0%	29.5%	9.2%	25.3%	7.2%	20.5%	10.4%	24.3%
Memphis	35.0%	18.2%	13.6%	25.6%	6.9%	27.2%	13.5%	25.4%
Nashville	32.7%	23.8%	10.1%	24.9%	7.5%	24.7%	10.4%	24.7%
Providence	28.9%	33.7%	11.8%	24.9%	7.9%	16.3%	11.8%	22.4%
Raleigh	32.4%	26.3%	12.0%	26.5%	8.0%	19.1%	12.2%	24.0%
Average	**30.1%**	**25.3%**	**11.2%**	**25.0%**	**7.5%**	**23.3%**	**11.8%**	**24.4%**

Source: 2001 National Household Travel Survey.

Besides survey data, the other main data source for validation of time-of-day models is traffic volume data by time of day. These data can be disaggregated for the periods used in the model and can provide estimates of total demand by time period. These data are useful for aggregate checks over all travel purposes (since the count data do not distinguish purposes). The most useful checks using traffic count data are those done following trip assignment. Transit ridership data provide a source for aggregate checks of the results of transit assignment by time of day.

8.2 Aggregate Checks

The main aggregate checks for time-of-day models include person trip/activity percentages, traffic volume checks, transit boarding checks, and activity duration checks.

Person Trip/Activity Percentages

Modeled percentages of person trips or activities by purpose and time period may be compared to observed percentages obtained from the expanded household survey data. This may be done by market segment to the extent segmentation is available in the model. For example, in all models, segmentation by geographic subarea is possible since the zones or parcels representing trip end locations can be aggregated. Geographic segmentation may use the same types of districts as used in trip distribution validation (see Section 6.2.2). Another possible segmentation variable is income level. However, fixed time-of-day factors are usually not estimated separately by market segment.

For activity-based models, this type of check is performed by tour or activity purpose and includes both primary activities and intermediate stops on tours. For tours, the checks should include comparisons of tour arrival times, departure times, and the combination of the two since they represent the alternatives in the model.

The logit time-of-day choice models used in activity-based models, which include a number of variables, may lend themselves more readily to validation and calibration by market segment. These models typically include person- and/or household-level variables that could be used for segmentation and can also be checked using geographic segmentation. It should be noted that since these models usually have more time periods, and for tours the combination of arrival and departure times must be validated, the observed data are stretched thinner. The error ranges associated with the observed percentages will therefore be greater than those for longer, more aggregate periods.

It is worth discussing the use of data from the Census Transportation Planning Package (CTPP) in time-of-day model validation. The CTPP data are relevant only for the home-based work purpose. However, the American Community Survey (ACS), on which the CTPP data are based, asks, “What time did this person usually leave home to go to work LAST WEEK?” This differs from what a time-of-day model for home-based work travel simulates. For example, if every person in a particular market segment left for work at

7:00 a.m. four out of five times and at 8:00 a.m. the fifth time, the CTPP would record answers of 7:00 a.m. for all respondents, but the model would simulate 80 percent of the travelers leaving at 7:00 a.m. and 20 percent leaving at 8:00 a.m. It is also important to note that ACS records only about trips to work, not home from work, the latter of which are also included in home-based work trips in four-step models. This also means that the return times for work tours are not reported, and so the data are insufficient for validating tour-based time-of-day choice models. The CTPP/ACS data also do not distinguish between trips made directly between home and work and those made with stops, which are often treated as more than one trip in four-step models. Another important point to consider is that the ACS records departure time while the time-of-day model may measure arrival or midpoint time.

National data sources such as NHTS should be used only as general guidelines for checks of time-of-day percentages for person trips. As the data in Table 8.1 showed, there is significant variation in temporal distributions of travel even among urban areas of similar size, and so calibrating factors for a specific area to match national distributions may be inaccurate.

Traffic Volume Checks

Following trip assignment, another source of data becomes available for validation, namely traffic count data. Often, traffic count data are available at an hourly (or finer) level, which is usually less aggregate than the periods used in highway assignment. It is usually not difficult, therefore, to aggregate count data to match the time periods used in highway assignment.

It is recommended that, even when highway assignment is done for separate time periods, the average daily assignment results (the sum of the assignment results for the various time periods) be validated and calibrated before any time-of-day validation is performed. Once the daily modeled volumes have been validated, the volumes by time period can be checked. While the values for the various guidelines may differ – for example, percentage root mean square errors would likely be higher for the lower peak period volumes than for daily volumes – the same basic checks performed for daily assignment results (see Section 9.1) are performed for each time period for which assignment is run.

Probably the first basic check for time-of-day assignments is total modeled vehicle miles traveled (VMT) by time period compared to observed VMT. Assuming the daily assignment results have already been validated, it is easy to compare the percentage of daily modeled VMT occurring during each period to the observed percentages. Other assignment checks that can be performed by time period include screenline volume checks and checks of volumes on key roadways.

Transit Boarding Checks

In a model with significant transit demand and a time-of-day model component, it is important to check the transit assignment results by time-of-day. This is done by comparing the modeled unlinked transit trips by time period to observed boardings by time period. Naturally, this assumes that transit ridership information by time-of-day is available. As is the case with highway trips, it is best to first validate the daily transit assignment (as described in Section 9.2) before validating the transit assignment by time period. The level of aggregation of transit routes for validation can be the same for time-of-day assignments as for daily assignments.

Activity Duration Checks

For activity-based models, the activity durations are implied by the modeled activity start and end times, or arrival and departure times. These implied activity durations can be compared to the actual activity durations from the household survey data on an aggregate basis. For example, the modeled versus observed activity durations can be compared by activity purpose, demographic segment, and/or geographic segment, depending on data availability. This can be done for both primary activities and activities occurring at intermediate stops on tours.

8.3 Disaggregate Checks

Disaggregate time-of-day model checks can be performed only for the logit time-of-day choice models usually associated with activity-based models. Fixed factor methods are aggregate and cannot be checked disaggregately.

As discussed in Section 6.2.3, logit models are disaggregately estimated, and therefore disaggregate validation is needed for logit time-of-day choice models, along with the aggregate checks described above. Most urban area household travel surveys have such small sample sizes that the entire data set is needed for model estimation. In most cases, therefore, there is no independent model estimation data set available for validation.

Limited disaggregate validation can be performed using the same data set used for model estimation, but reporting the results by market segment, using the "apply" function in the logit model estimation software. A logit time-of-day choice model could be applied to the data set used for estimation but the results may be reported by, for example, worker status. It might be found, for example, that the model inaccurately estimates peak period shopping trips for part-time workers.

8.4 Criteria Guidelines

While it is not difficult to statistically estimate error ranges or confidence intervals for time-of-day percentages, these values would vary significantly based on survey sample sizes and trip/activity purpose definitions. It would be possible, though, to compute such intervals and determine whether modeled time-of-day percentages fall within an acceptable range (say a 90- percent confidence interval). This has not been done often in practice, perhaps in part because in many cases the confidence intervals would be so large as to be essentially meaningless. A similar issue arises for the comparison of modeled time percentages from assignment results to traffic or transit ridership counts. The error ranges for an individual traffic count can be quite high compared to the value of the count itself.

Because there is no standard practice, no specific criteria guidelines for the aggregate checks described in Section 8.3 are presented. The modeler must judge whether the modeled percentages are "close enough," based on the statistical validity of the observed data and the need for accuracy in model results.

8.5 Reasonableness and Sensitivity Testing

Some reasonableness checks for time-of-day models can be done. It is logical for home-based work trips to have pronounced morning and evening peaks and for nonhome-based trips to have a mid-day peak. When time-of-day factors or models are applied, the resulting split of home-based production-to-attraction trips and attraction-to-production trips should be even. It might be reasonable, however, for the trips for individual home-based trip purposes to differ from a 50-50 split between directions. For example, in most cases, there are more stops made on the way home from work, implying that there are likely more production-to-attraction trips than attraction-to-production trips. However, this should be offset by a greater number of attraction-to-production trips for other home-based trip purposes.

Sensitivity testing is not relevant for fixed factor models, which, by their use of fixed factors, are insensitive to variables related to transportation level of service or household or person demographics. Sensitivity testing can be performed for time-of-day choice models by varying model inputs and checking results for reasonableness. Model inputs that can be varied include level of service variables used in the model (e.g., distance, delay) and demographic variables. Some example tests include:

- Increasing or decreasing travel times/delays for a specific time period or periods by a fixed percentage regionwide to determine the change in peak demand (this would represent "peak spreading" in the case of increased peak travel times); and
- Reallocating the number of households or person by a specific input variable for a forecast year.

Due to a lack of experience, developing specific criteria guidelines for expected changes in temporal distribution patterns based on changes in input variables is difficult. Once again, it is up to the experienced planner to determine whether the modeled changes are reasonable.

8.6 Troubleshooting Strategies

Table 8.2 summarizes some of the issues that might be discovered during the validation checks described above and potential strategies to address them.

Table 8.2 Troubleshooting Strategies for Issues with Time-of-Day Model Results

Issue	Potential Troubleshooting Strategies
1. Modeled percentages of trips by purpose and time period poorly match observed survey data	• For fixed factor models, recheck factor application in the model • Check for consistency between application data (trip tables) and estimation data (survey) • For time-of-day choice models, recalibrate time period constants
2. Modeled percentages of assigned trips by time period poorly match independent observed data (traffic/transit counts)	• For fixed factor models, consider revising factors to achieve a better match • For time-of-day choice models, if issues are regionwide, consider recalibrating time period constants or other parameters • For time-of-day choice models, if issues are mode-specific, consider revisions to mode-specific parameters
3. Model results by geography or socioeconomic strata differ from percentages based on expanded household survey data	• For fixed factor models, consider whether factors could be varied by geography or other appropriate market segment • For time-of-day choice models, consider adding segmentation variables or calibrating coefficients for existing variables

8.7 Forecasting Checks

The basic forecasting checks for time-of-day models are comparisons between percentages of travel by time of day between the base and forecast years. All of the checks discussed in Section 8.2 can be performed comparing the base and forecast year scenarios.

The comparisons are not likely to show much useful information for fixed factor models. The percentages of person trips by purpose for each time period are fixed and therefore must be equal for all scenarios. The percentages of total travel across purposes, including person trips and assigned vehicle and transit trips, may vary between the base and forecast years since the distribution of daily trips among trip purposes may differ, due to land use and demographic changes. These differences in the temporal distribution of overall travel can be checked for reasonableness, but the differences are likely to be small.

A potential problem in areas of high growth is that forecasted peak demand may exceed capacity in some locations. Since fixed factor models do not provide a direct way to consider peak spreading, the highway assignment may result in some roadways being assigned volumes well over capacity. If these results are too unrealistic, it may be necessary to address peak spreading using a post processor that reallocates trips from peak to off-peak periods. It is important, however, to consider that unrealistically high levels of congestion in forecasts may be due to other issues with the model, such as network coding problems, highway assignment parameters, or issues with feedback procedures.

For time-of-day choice models, the aggregate checks discussed above (for tours and trips as appropriate), plus checks of activity durations, should be performed comparing the base and forecast years. Because these models should be sensitive to levels of congestion, peak spreading may be revealed by these comparisons. A good reasonableness check would be to see if peak spreading has occurred in very congested locations, and that little change has occurred in uncongested areas. When doing these checks, however, care should be taken to consider changes in temporal distributions that may be due to other variables, such as person and household demographics. Perhaps the simplest way to consider the results of the comparisons for time-of-day choice models would be to discover an explanation for significant changes in temporal distribution, whether it is due to changes in demographics, levels of congestion, or other variables.

9.0 Assignment Procedures

9.0 Assignment Procedures

Assignment is often viewed as the culmination of any modeling process, be it a traditional four-step modeling process or an activity-based modeling process. Many models now include feedback loops to "equilibrate" assigned travel speeds with travel speeds used for prior modeling steps such as trip distribution, destination choice, and mode choice. Nevertheless, the modeling process typically ends with the assignment step.

The assignment step includes both highway and transit assignments of vehicle and person trips respectively. While there are emerging assignment procedures such as dynamic traffic assignment (DTA) and regional simulation procedures, research into the integration of these emerging procedures and travel demand models is just now occurring.

Assignment validation is generally inseparable from the rest of the modeling process. This is especially true for traffic assignment since it is not feasible to collect sufficient survey data to construct an observed trip table for traffic assignment. For transit assignment, observed transit trip tables might be constructed from comprehensive on-board surveys such as those performed for the FTA New Starts analyses.[31]

Assignment validation is an important step in validating not only the assignment process but the entire modeling process. Assignment validation typically benefits from a wealth of independent validation data including traffic counts and transit boardings collected independently of household or other survey data used for model estimation and, increasingly, from independent traffic speed and travel time studies. In addition, due to established traffic and transit counting programs in many regions, traffic and transit count data can be used for temporal validation of travel models (see Chapter 10).

Unfortunately, as the culmination of the modeling process and due to the wealth of independent validation data, the assignment of trips to the network often becomes the primary basis for validating the a travel model's ability to replicate observed travel. In effect, assignment validation becomes a "super" data point defining a successful validation for many modelers and planners. While it is important that assignment validation be reasonable, highly accurate traffic and transit assignments in terms of matching observed traffic and transit volumes are *not* sufficient for proving the validity of travel models. In some cases, the over-emphasis on matching observed traffic volumes and transit boardings has led to poor model adjustments such as link specific changes to the network speeds and

[31] The use of observed transit trip tables for transit assignment validation is covered in this chapter. However, it is also important for validating transit networks (see Chapter 3) and for inputs for validating trip distribution (Chapter 6) and mode choice (Chapter 7).

capacities and "fine-tuning" of connector links for better match between modeled and observed traffic volumes or transit boardings.

Since assignment techniques are not wedded to a specific modeling process, this chapter will be structure slightly differently from the other chapters in this manual. Specifically, it will focus first on traffic assignment validation and then on transit assignment validation.

9.1 Traffic Assignment Checks

Both traditional and emerging traffic assignment procedures may be used for assignment. Traditional techniques may be characterized as procedures that represent trips on each interchange as being omnipresent on all links reasonably serving the interchange. Traditional techniques include static equilibrium assignment, other capacity-restrained assignment, stochastic multipath assignment, and all-or-nothing assignment. Static equilibrium assignment is probably the most frequently used traditional traffic assignment technique.

Capacity-restrained traffic assignment techniques rely on volume-delay functions to estimate increases in individual link traversal times as assigned traffic volumes approach the estimated traffic carrying capacity for the link. The Bureau of Public Roads (BPR) curve has often been used to estimate link travel times resulting from the assigned volumes. In recent years, a number of enhancements have been made to the process, due in part to increases in computing power. Volume-delay functions have been developed for different facility types (freeway versus arterial for example) and in some regions, intersection-based techniques mimicking Highway Capacity Manual (HCM) intersection delay estimation techniques have been implemented. The detail of the coding of networks has also increased dramatically, along with an associated reduction in the size of the traffic analysis zones. Most traffic assignment programs also provide an option for class-based traffic assignment so that single-occupant vehicles (SOV), high-occupant vehicles (HOV), and trucks can be assigned simultaneously, interacting on general purpose links, but also being able to travel on links restricted by vehicle class (e.g., HOV lanes).

Emerging traffic assignment techniques include DTA and regional traffic simulation. A key to the emerging techniques is that they explicitly account for the actual time to travel between an origin and destination for an interchange. In addition, the emerging techniques can account for traffic queues backing up to impact other links in the network. The emerging traffic assignment techniques may be more suitable for use with activity-based modeling techniques although some have been applied using the results of traditional four-step models.

The focus of this section is the validation of traffic assignments. Many of the validation techniques relate to link-based traffic volumes and travel times. The validation tests can be applied regardless of whether the assignment results were produced by a traditional assignment technique or an emerging technique.

9.1.1 Sources of Data

Traffic Counts

Traffic count data are the primary data used for the validation of traffic assignment procedures. Most traffic count data are obtained from various traffic count programs used for monitoring of traffic or collected for the Highway Performance Monitoring System (HPMS).

Traffic count data are an important independent validation data set. Nevertheless, traffic count data are often afforded more credence than they deserve. Counts are often collected from multiple sources such as state Departments of Transportation, toll authorities, counties, cities, and private contractors with each using various counting techniques. For example, counts from permanent traffic recorders, 48-hour or 24-hour counts performed using tube counters, and ancillary counts such as manual intersection counts may all be stored in the same database. Counts may be stored as raw counts or factored counts, such as average annual daily traffic (AADT). In addition, counts from multiple years surrounding a base year for model validation may be included for a validation in order to maximize the count data available.

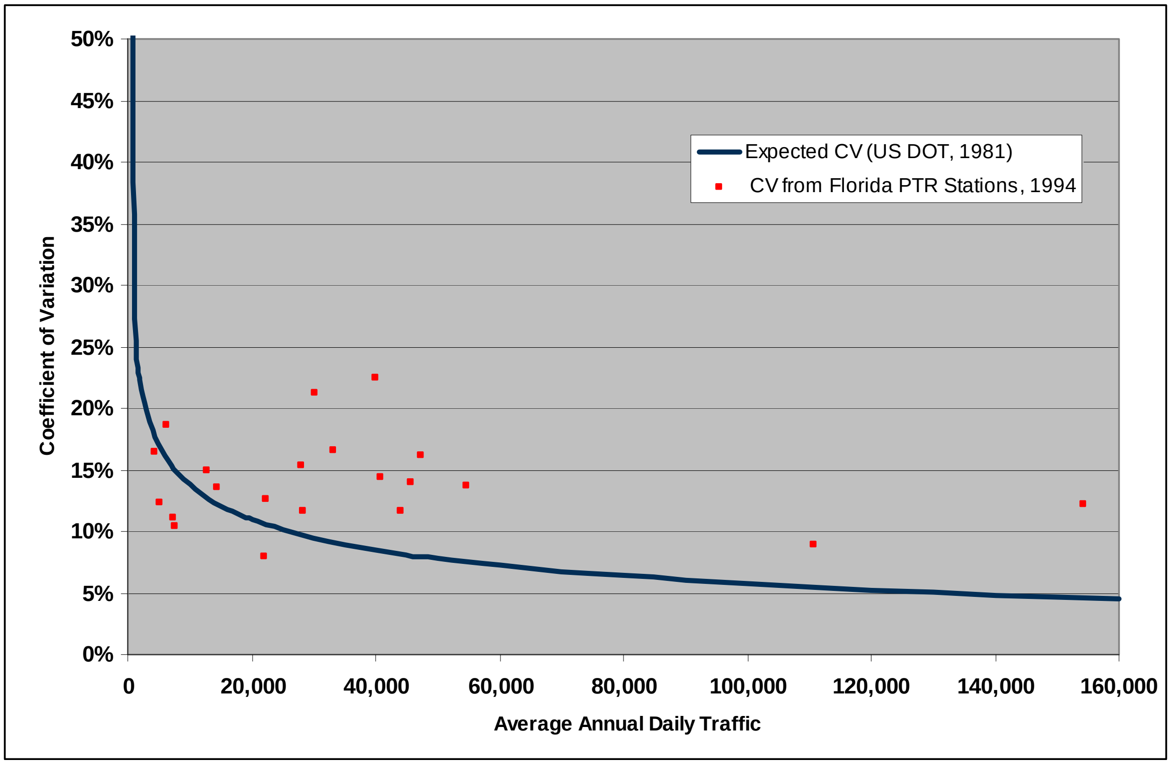
In light of the above, the development of a validation database is a significant undertaking. In establishing the database, the data forecast by the regional travel model should be considered. Most regions develop travel models to provide forecasts of travel for an average weekday. Thus, the traffic count validation data should also reflect average weekday traffic (AWDT) for consistency. In addition to ensuring consistency of counts, the development of the traffic count database should also include consideration of geographic coverage, adequate representation of different functional classes, and completeness of screenlines. Inclusion of classification count data should be considered, especially if the travel model produces (and the region is concerned with) forecasts of high occupancy vehicles or truck volumes.

The variation of the count data should also be a concern in the development of the traffic count validation database. A traffic count for a facility is, in effect, a single sample of the set of daily traffic counts that occur on the link over a period of time. Thus, a single traffic count or a set of traffic counts for a single facility represent a sample for the link subject to sampling error. In 1981, the U.S. Department of Transportation published the Guide to Urban Traffic Counting, which included a figure depicting the expected coefficient of variation in daily counts. In 1997, a study of the variability of traffic count data included information from 21 permanent traffic recording (PTR) stations in Florida.[32] The curve depicting the original estimation of coefficient in variation of traffic counts and the observed data from Florida are shown in Figure 9.1.

---

[32]Wright, Tommy, et al., Variability in Traffic Monitoring Data, Final Summary Report, prepared for Oak Ridge National Laboratory, August 1997.

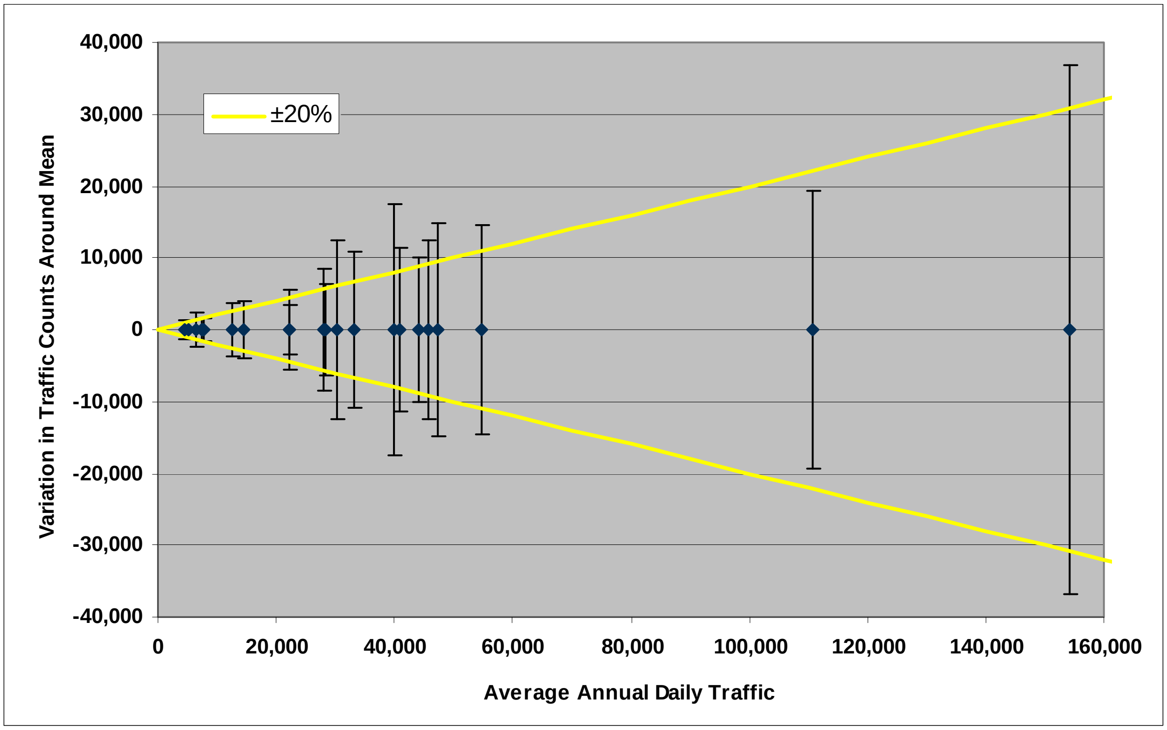
Figure 9.1 Expected Coefficient of Variation in Daily Count Volume and Observed Coefficients from Florida Permanent Traffic Recorders



Sources: U.S. Department of Transportation, Guide to Urban Traffic Counting, 1981; and Wright, Tommy, et al., *Variability in Traffic Monitoring Data, Final Summary Report*, prepared for Oak Ridge National Laboratory, August 1997, Table 5, Page 10.

Since coefficient of variation is defined as the standard deviation divided by the mean, the standard deviation of an observed traffic volume can be easily estimated. Approximately 95 percent of the randomly collected counts for a facility should fall within ±1.96 standard deviations of the mean. Figure 9.2 shows the error ranges (±1.96 standard deviations) for the Florida data along with lines representing ±20 percent of the average traffic count. As can be seen in Figure 9.2, substantial variation in daily traffic counts should be assumed. This suggests that traffic count data based on one or two day counts may be substantially different than the "true" average daily traffic for a link, even when the traffic count data are adjusted for day of week and seasonal variation. These results suggest that link counts for subclassifications such as time-of-day or vehicle classification are also subject to substantial variation.

Figure 9.2 Variation in Observed Traffic Counts
Florida Permanent Traffic Recorder Data



Source: Wright, Tommy, et al., *Variability in Traffic Monitoring Data, Final Summary Report*, prepared for Oak Ridge National Laboratory, August 1997, Table 5, Page 10.

HPMS Data

Regional vehicle-miles of travel (VMT) are estimated from traffic counts for the HPMS. The regional VMT estimates provide a target for modeled VMT. However, prior to using the observed regional VMT based on the HPMS data, the consistency of the HPMS data and the modeled data should be verified. The consistency checks should include:

- The HPMS area covered versus area covered by the travel model;
- The facilities included in HPMS VMT (e.g., local street VMT) versus facilities included in model; and
- Whether VMT estimates are based on average annual daily traffic or average annual weekday traffic.

Travel Time and Speed Studies

Many regions have initiated the collection of travel times and speeds. These studies typically identify a number of corridors in the region each served by a single functional class such as freeway, expressway, principal arterial, or minor arterial. A number of travel time runs are then made through the specified corridors at various times of day to collect travel time, and thus, average travel speed information. The data collected can vary from simple end-to-end travel times to the components of the end-to-end travel times including run times, cruise times and signal delay times, delay times due to incidents, and in some studies, coincident traffic counts on the facilities traversed.

If traffic count data are collected along with the detailed travel time data, it may be possible to use the data to validate (or even to estimate) the volume-delay functions used in the traffic assignment process. Some regions have used detailed travel time and traffic count data to develop volume delay functions that result in validated traffic counts and traffic speeds being produced directly by the assignment process. Other regions use one set of volume-delay functions to produce validated traffic counts and a second set in an assignment post-processing step to estimate traffic speeds for air quality modeling.

As with traffic count data, travel time and speed studies may be subject to substantial variation depending on the day or days the data are collected. Nevertheless, the data collected can be quite useful in validating congested speeds produced by the travel model. With the strong connection between travel models and air quality models, the validation of congested speeds produced by the traffic assignment procedure is an important consideration.

Some regions also collect spot speed study data. These data may be useful for validation of modeled speeds for facilities uninterrupted by intersections such as freeways and expressways. Spot speed data are of limited use for arterials and other facilities with traffic control devices at intersections since delays resulting from the traffic control devices are not considered in the speed studies.

9.1.2 Aggregate Checks

A good approach to the validation of the traffic assignment procedure is to start with the most general aggregate checks and progress toward more detail. Aggregate checks should be generally applicable for both traditional traffic assignment procedures and for emerging techniques.

Vehicle-Miles Of Travel

Base year VMT produced by the model can be compared to observed VMT estimated from the traffic count data and from HPMS data. When comparing to VMT estimated from traffic count data, modeled traffic for only those network links with traffic counts should be included in the estimation. For comparisons with HPMS VMT estimates, modeled traffic for all network links should be considered.

The VMT checks should be made for the region and by market segment. Markets may include facility type, area type, geographic subdivision (e.g., county or super-district), time-of-day (e.g., morning peak period, afternoon peak period, mid-day and night). Table 9.1 provides an example table for VMT summaries by facility type.

Table 9.1 Example VMT Validation Summary by Facility Type

	VMT		Error		Distribution	
Facility Type	**Estimated**[a]	**Observed**[b]	**Difference**	**Percent**	**Estimated**	**Observed**
Freeways	23,342,838	24,078,537	-735,699	-3%	37%	39%
Expressways	3,477,618	3,306,422	171,196	5%	6%	5%
Principal Arterials	19,508,011	18,578,391	929,620	5%	31%	30%
Minor Arterials	7,125,530	7,257,875	-132,345	-2%	11%	12%
Collectors	8,911,433	9,178,980	-267,547	-3%	14%	15%
Total	**62,365,430**	**62,400,204**	**-34,774**	**0%**	**100%**	**100%**

[a] Estimated is the VMT produced by the model.

[b] Observed is based on either traffic counts or the HPMS estimates of VMT.

As mentioned previously, assignment is the culmination of the modeling process and, in effect, validates the entire modeling process. The VMT checks provide this overall modeling process check more than subsequent tests that will be described later in this chapter. Different information regarding the modeling process can be inferred from each level of the summaries:

- Regional VMT summaries provide an indication of the reasonableness of the overall level of travel. The results help confirm that the trip generation, trip distribution, and mode choice models, or their activity-based modeling corollaries, as well as the assignment process, are performing reasonably.

- VMT summaries by facility type provide an overall indication of the operation of the assignment procedures. These results of these summaries might indicate issues with free-flow speeds, link capacities, or volume-delay functions.

- VMT summaries by geographic area may be useful for uncovering geographic biases in the modeling process. These biases might relate to previous steps in the modeling process. GIS plots of errors or percent errors by geographic area may facilitate this analysis.

- VMT summaries by combinations of the above strata may provide additional diagnostic information if one of the above summaries indicates a validation problem.

Traffic Volume-Related Checks

Traffic volume related checks compare modeled to observed traffic volumes on a link-by-link basis. Consequently, the amount of difference between the modeled and observed traffic for each link contributes directly to the overall measure of closeness even when the results are aggregated in different ways. This is in contrast to the VMT checks described above where a positive difference on one link can cancel a negative difference on another link.

The traffic volume related checks described in this chapter focus on traditional measures that are scalable and easily explained: root mean squared error (RMSE), percent RMSE (%RMSE), correlation (R), and coefficient of determination (R^2). There are other measures similar to the measures covered in this section, such as mean absolute error (MAE), that may be used or preferred by some. The key to the measures is that they are scalable. For example, an RMSE of 1000 is one-half as large as an RMSE of 2000 for a given set of links.

"Pass-fail" validation tests are not recommended or discussed in this section since they imply an unwarranted level of confidence in the results (and in the observed data) and do not provide useful information regarding the goodness of fit of the model. These measures can be characterized as "the results are 'valid' if the value obtained for the validation test is less than five."

Root Mean Squared Error and Percent Root Mean Squared Error

RMSE and %RMSE for a set of links can be calculated using the following formulae:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left[(\text{Count}_i - \text{Model}_i)^2\right]}{N}}$$

and

$$\%\text{RMSE} = \frac{\text{RMSE}}{\left(\frac{\sum_{i=1}^{N}\text{Count}_i}{N}\right)} \times 100$$

Where:

$Count_i$ = The observed traffic count for link *i*;

$Model_i$ = The modeled traffic volume for link *i*; and

N = The number of links[33] in the group of links including link *i*.

RMSE and %RMSE are both measures of accuracy of the traffic assignment measuring the average error between the observed and modeled traffic volumes on links with traffic counts. As such, RMSE and %RMSE should be summarized by facility type (or functional class) or by link volume group. Summarizing the measures by geography can provide good validation information, especially if the measures continue to be stratified by facility type or volume group. While the measures can be calculated for more aggregate groups or the region as a whole, the measure becomes less useful for determining the quality of the assignment process. In effect, at too gross a level of aggregation, the RMSE or %RMSE measures can easily be interpreted as pass-fail measures: "The regional %RMSE is 32 percent so, obviously, the model is…" Such statements have little validity or usefulness for model validation.

If the traffic assignment process used for a region uses a look-up table to estimate link capacity (e.g., stratified by area type and facility type), it is useful to summarize RMSE by the same strata. In this way, the average error on links can be compared to the estimated capacities of the links to determine if the average error is, say, more or less than one-half lane of capacity. If the RMSE is based on more than a one-hour assignment, as would typically be the case, the RMSE can be adjusted to reflect a one-hour period through the use of a peak hour factor. For example, suppose the RMSE for freeways in a suburban area type was 10,000 based on daily traffic counts and the modeled daily traffic volumes. If eight percent of the daily traffic occurred in the peak hour, the average error represented in the peak hour could be estimated as 0.08 x 10,000, or 800 vehicles. If the modeled capacity for freeway links in the suburban area was 2,200 vehicles per hour per lane, the implied average error would be equivalent to a little over one-third of a lane.

Correlation Coefficient or Coefficient of Determination

Pearson's product-moment correlation coefficient (R) is a standard statistical measure available in spreadsheet programs and other readily available statistical software packages. R is a dimensionless index that ranges from -1.0 to 1.0 inclusive that reflects the extent of a linear relationship between two data sets. It is calculated as follows:

---

[33]Some analysts prefer using "(N-1)" in the denominator of the RMSE calculation for an unbiased estimate of RMSE. Practically, this adjustment to the equation has decreasing impact as the number of observations increases. With N=10 links, the "unbiased" formula increases the value of RMSE about five percent over the formula shown above. With N=50, the unbiased RMSE is about one percent higher than the above formula, and with N=100, the unbiased RMSE is about 0.5 percent higher. The form of the RMSE equation shown above is consistent with the 1997 version of this manual. The choice of the formula is up to the analyst.

$$R = \frac{N \times \left[\sum_{i=1}^{N}\left(\text{Count}_i \times \text{Model}_i\right)\right] - \left(\sum_{i=1}^{N}\text{Count}_i\right) \times \left(\sum_{i=1}^{N}\text{Model}_i\right)}{\sqrt{\left[\left(N \times \sum_{i=1}^{N}\text{Count}_i^2\right) - \left(\sum_{i=1}^{N}\text{Count}_i\right)^2\right] \times \left[\left(N \times \sum_{i=1}^{N}\text{Model}_i^2\right) - \left(\sum_{i=1}^{N}\text{Model}_i\right)^2\right]}}$$

Where:

Count_i, Model_i, and N are as defined for the calculation of RMSE.

The coefficient of determination, R^2, which is simply the square of R, is typically interpreted as the proportion of the variance in a dependent variable, y, attributable to the variance in an independent variable, x. This traditional interpretation does not hold for traffic assignment validation since the modeled traffic assignment is not dependent on the traffic count, or vice-versa.

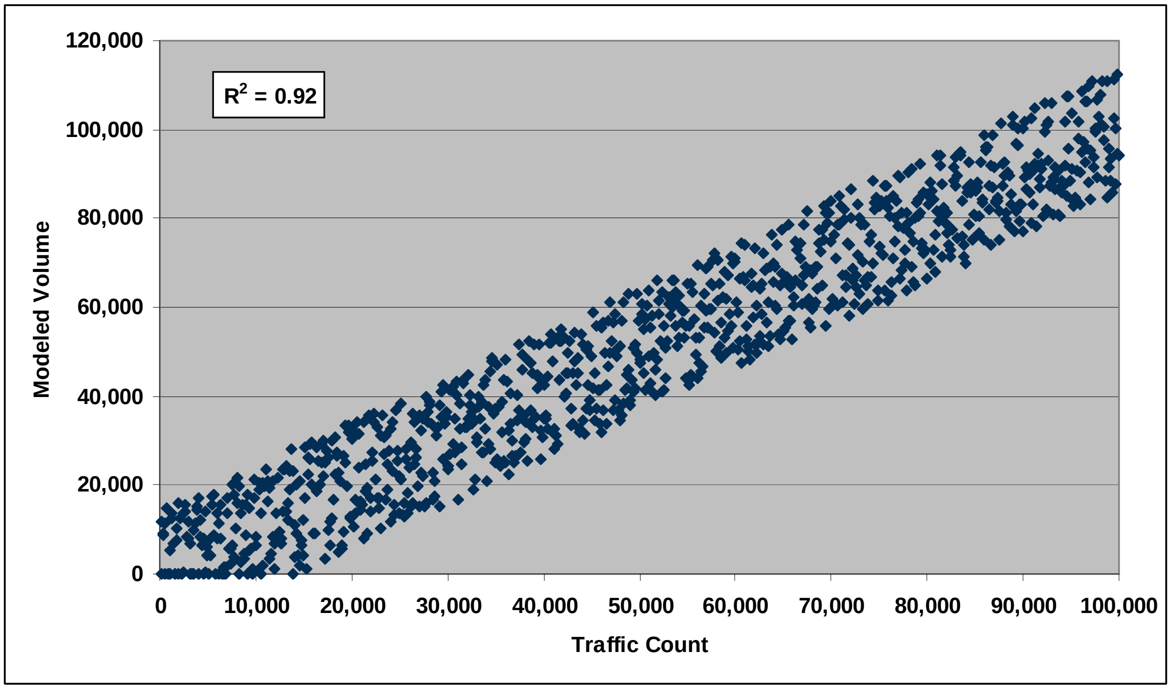
These two measures have been frequently used in the past in validation. They measure the strength of the (linear) relationship between the assigned volumes and traffic counts. In effect, R^2 has been assumed to be a measure of the amount of variation in traffic counts "explained" by the model. The measures must be used with caution. An R^2 for all links in the region simply says that links with high capacities (e.g., freeways) can, and usually do, carry more traffic than links with low capacities (e.g., local streets). As such, R^2 probably tells more about the coding of facility type and number of lanes than about how the model and assignment is performing. Thus, achieving a regional R^2 of 0.88, as has been suggested as a "standard" for determining a model's validity, has little if any meaning.

If used carefully, R^2 can be a useful measure for comparing model results to other iterations when calibrating travel models and traffic assignments since the bases (i.e., the sets of links considered) for calculating the measure should be the same between iterations. The R^2 statistics should be calculated for links with similar characteristics such as facility type or volume group. As an example, if the R^2 statistics for each facility type were consistently higher for Iteration "X" of a travel model calibration as compared to the results for other iterations, the model used for Iteration X *might* be considered to be the best. Of course, all modifications made to the model for Iteration X should be considered prior to ranking the final results of the various iterations.

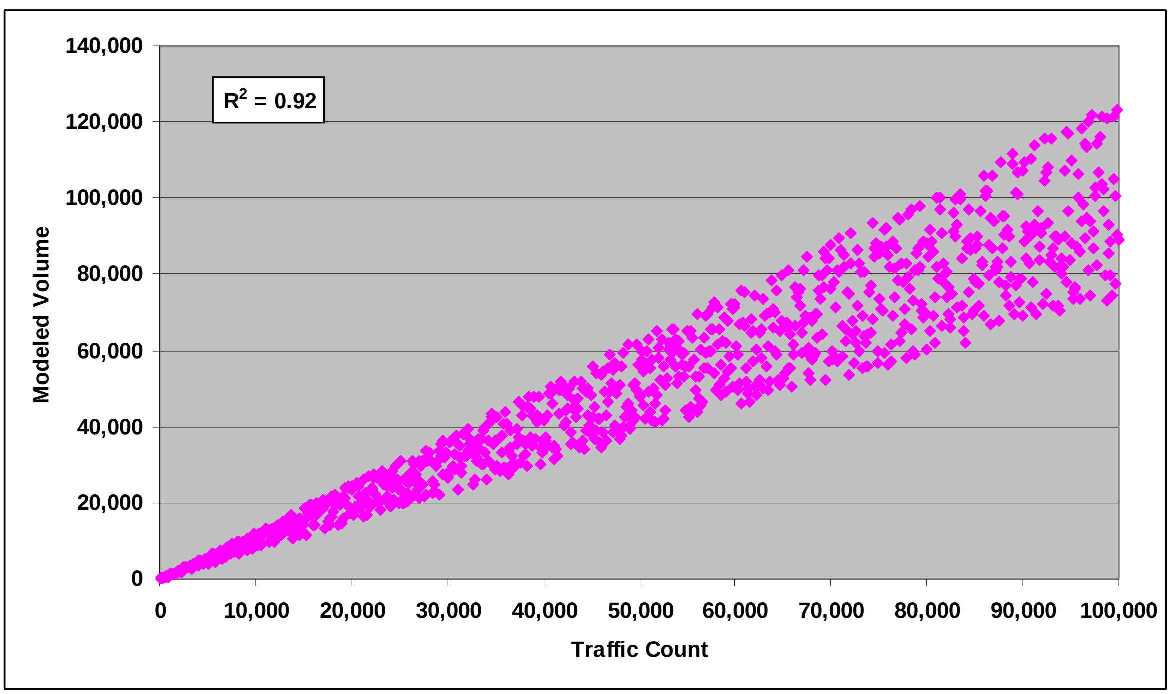
Scatterplots

Scatterplots of modeled traffic volumes versus the observed traffic volumes are useful validation tools and should be combined with the R^2 summaries. Figure 9.3 shows two scatterplots with identical R^2 values. Even though the R^2 values are identical, the scatterplots tell very different stories regarding the modeled volumes. In Figure 9.3(a), the modeled volumes are randomly distributed around the observed traffic counts within a constant band. Such results might suggest that the volume-delay functions are having relatively little effect in the traffic assignment. In Figure 9.3(b), the scatterplots suggest that the amount of error in the modeled volumes is proportional to the traffic count or, in effect, to the capacity of the link.

Figure 9.3 Example Scatterplots of Modeled Traffic Volumes Versus Observed Traffic Counts



(a)



(b)

Range Checks

Analysis of outliers can be a good method for finding and correcting network or assignment errors. Some outliers, links with high observed volumes and very low assigned volumes or vice-versa, can be identified from the scatterplots. An alternative method for identifying outliers is to simply list or plot the links with the largest differences between modeled and observed traffic volumes. It is also worthwhile to identify and investigate links with zero assigned volumes.

Screenlines, Cutlines, and Cordon Counts

Comparison of modeled volumes to observed counts for critical links, especially along screenlines, cutlines, and cordon lines, are useful for assessing model quality:

- Screenlines extend completely across the modeled area from boundary cordon to boundary cordon. Screenlines are often associated with physical barriers such as rivers or railroads, although jurisdictional boundaries such as county lines that extend through the study area may also be used as screenlines. Figure 9.4 shows example screenlines for a region.

- Cutlines extend across a corridor containing multiple facilities. They should be used to intercept travel along only one axis. Figure 9.5 shows example cutlines for multiple corridors a region. Cutlines 3, 6, 7, and 8 might be also considered screenlines if the entire modeling area is shown in Figure 9.5.

Figure 9.4 Example Screenlines

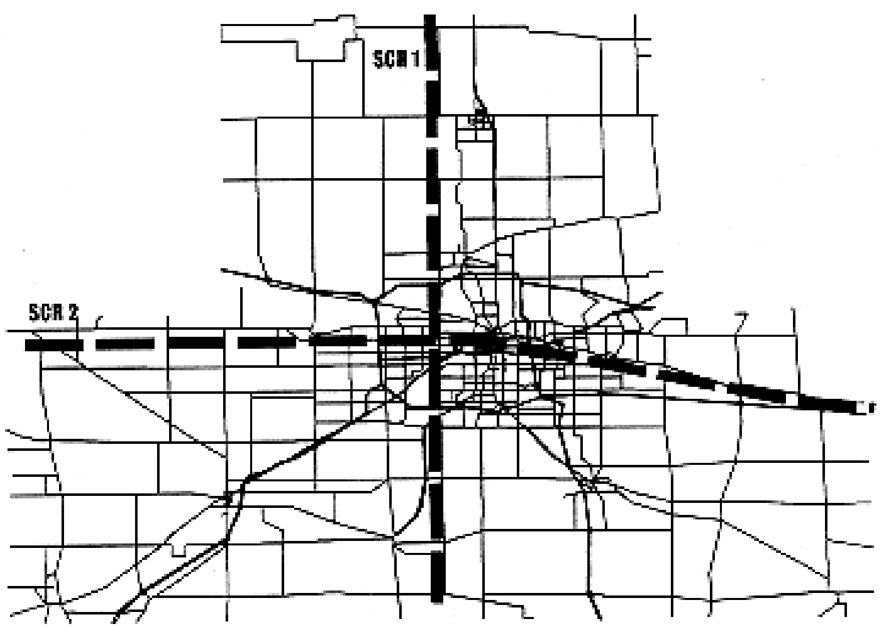
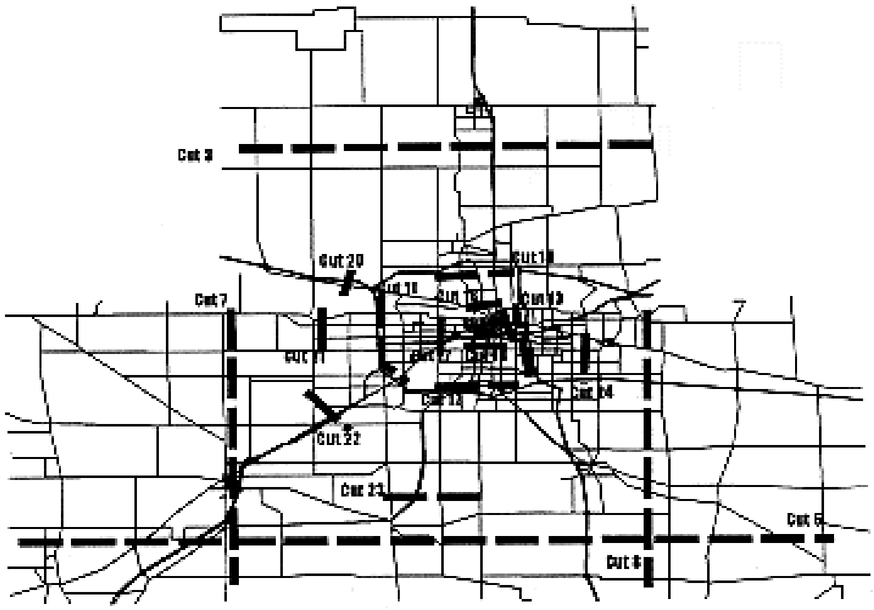


Figure 9.5 Example Cutlines



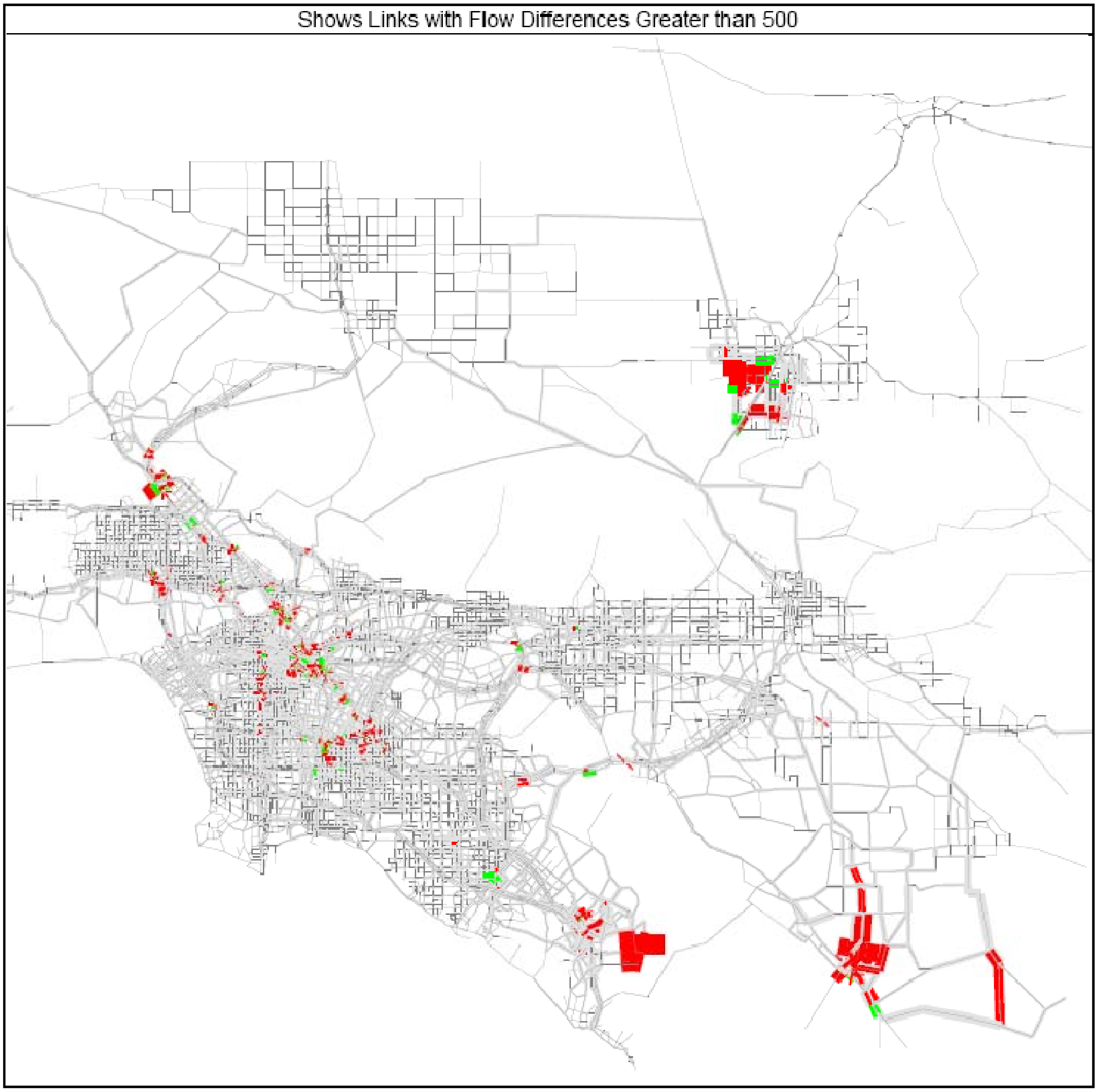
- Cordon lines completely encompass a designated area. For example, a cordon around the central business district is useful in validating the "ins and outs" of the CBD related traffic demand. Over or under estimates of trips bound for the CBD could indicate errors in the socioeconomic data (employment data for the CBD) or errors in the trip distribution or mode choice model.

Detailed Difference Plots

Detailed plots of absolute or relative differences between modeled traffic volumes and observed traffic counts can provide useful diagnostic information for model validation. Figure 9.6 shows an example of such a difference plot. Detailed difference plots are more appropriate for validation of models for corridor studies or diagnosis of problems. Typically, there is too much information at a regional level, although the data may be filtered to show only differences greater than a specified threshold value.

Since difference plots are comparing modeled volumes and observed counts on individual links, care should be exercised in the review. As noted in Section 9.1.1, there can be substantial "sampling" error in traffic counts. Thus, a large difference between modeled traffic volumes and observed traffic counts might reflect a problem with the traffic count rather than the modeled volume. The other aggregate tests described in this chapter are somewhat dependent on the sampling error associated with traffic counting being distributed around a reasonable mean. In other words, the other aggregate tests are dependent on over-counts for links of a specific group canceling under-counts for the group.

Figure 9.6 Comparison of Assignment Results



Speed Checks

Speed checks compare modeled speeds to observed data from travel time studies or, possibly, spot speed data for facilities not affected by intersection controls. The modeled speeds may be output directly from the traffic assignment process or they may be output from an assignment post-processor. The speed checks are focused on time-of-day or peak hour assignment results. While they can be easily calculated from VMT and vehicle-hours of travel (VHT) summaries for links, 24-hour average speeds are not very meaningful.

It is somewhat more difficult to define validation tests focused on speeds than it is to define traffic volume related validation checks. While modeled speeds can easily be calculated for each link, the modeled speeds are directly impacted by the quality of the assignment results. Thus, errors in assigned speeds might result from errors in the estimation of speeds or from errors in assigned traffic volumes. This issue might be addressed by filtering the links included in the test to include only those links where the assigned traffic volume is within, say, ±20 percent of the observed traffic count.

Scatterplots

An initial validation check of modeled speeds can be prepared by producing scatterplots of modeled versus observed speeds. The scatterplots might look like the examples shown in Figure 9.3 with "Observed Speed" and "Modeled Speed" replacing "Traffic Count" and "Modeled Volume." The scatterplots should be produced by facility type and, if possible, by link volume group within the facility type grouping. The stratification by volume group would address two primary issues:

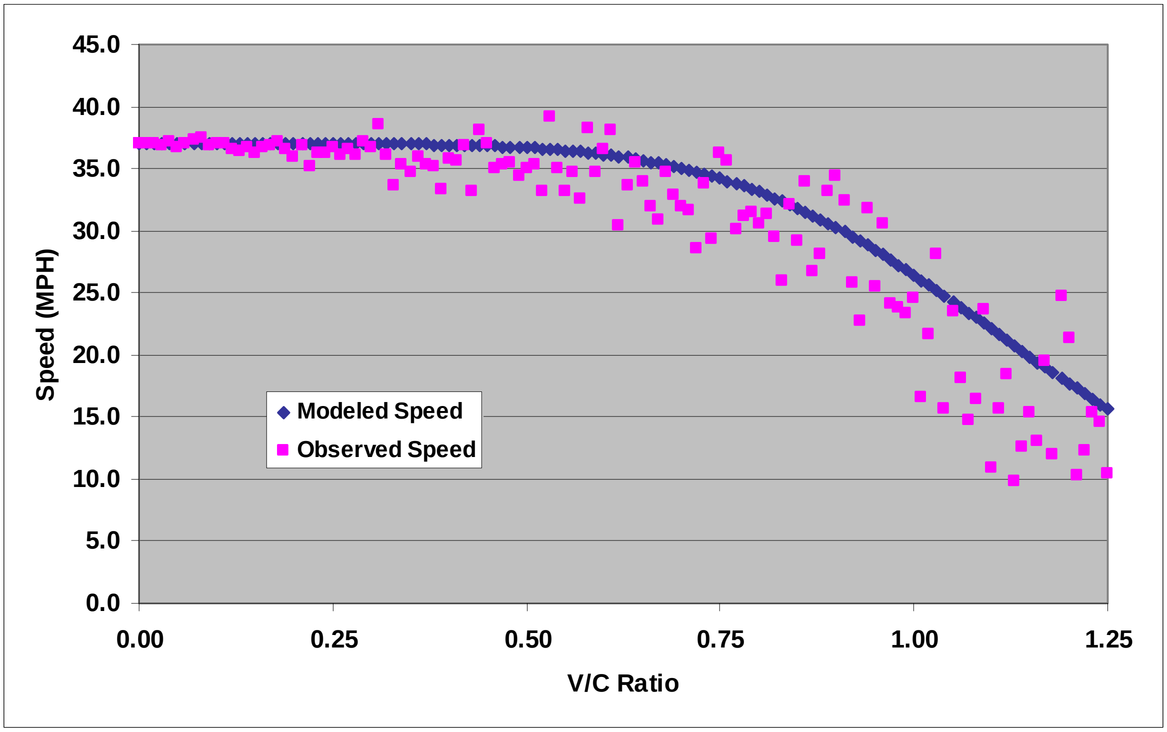
- It is probably more desirable to match traffic speeds on high volume links than on low volume links; and
- Speeds on low volume links should be close to free-flow speeds; if the free-flow speeds do not match reasonably, the veracity of the volume delay functions or the free-flow speed inputs can be questioned especially if the speeds for high volume links match closely.

Speed Versus Volume/Capacity Ratio Comparison Plots

Both observed and modeled speeds can be plotted against volume/capacity ratios. The observed speeds should be plotted against the volume/capacity ratio for the observed traffic count at the time the speed information was collected. The modeled speeds should be plotted against the modeled volume/capacity ratio. The plots should be produced by facility type. Figure 9.7 shows an example of such a plot.

The comparison plot shown in Figure 9.7 is a method for verifying volume delay functions for the assignment. It is just as valid to plot the modeled speeds using the specified volume-delay function for a specified facility type. The comparison plots remove the impacts of differences in modeled traffic volumes and observed traffic counts inherent in the scatterplots of modeled versus observed speeds. The plot shown if Figure 9.7 suggests that the modeled speeds do not decrease quite quickly enough as the volume/capacity ratio increases.

Figure 9.7 Example Comparison Plot of Speeds versus Volume/Capacity Ratios



Travel Time Run Comparison

This test can be implemented when speed and travel time surveys (i.e., "speed runs") have been performed for specific corridors. The test is implemented by comparing the modeled travel times over a specific route to the observed travel times for the same route. The test provides a general assessment of the overall quality of the traffic assignment and speed processing. Since slow speeds on some links in the route considered can cancel high speeds for other links, the test is not a rigorous test. Nevertheless, general trends can be observed if there are multiple speed runs for the same or for different corridors.

With the increased use of global positioning system (GPS) units for household travel surveys, there will be an increase in "speed run" data for model validation. Since one of the assumptions underlying static-based equilibrium traffic assignment is that no traveler can reduce his or her travel time by switch travel paths, it should be increasingly possible to compare travel times on interchanges for selected times of day to modeled travel times for the same interchanges for comparable time periods. These comparisons will provide general information regarding the reasonableness of modeled travel speeds.

9.1.3 Disaggregate Checks

Disaggregate validation checks are focused on emerging traffic assignment techniques such as DTA and traffic simulation. As such, validation methods are also emerging and may require data that are not readily available. The following outlines two possible tests for the emerging techniques.

Route Choice

If household or other travel survey data have been collected using GPS units, it might be possible to compare modeled to observed paths for selected trips. A measure of accuracy such as the percent of modeled links used matching observed links used might be useful.

Traffic Flow

Aggregate tests such as link speed comparisons and traffic volume comparisons described above are useful for validation of the emerging techniques. However, additional tests might be appropriate, especially if GPS data are available. Specifically, for specific trips, it might be possible to compare the components of travel time for a selected route (e.g., stop delay time and travel time in motion). Alternatively, if traffic engineering data are available, modeled level of service measures (e.g., intersection delay) might be compared to observed data.

9.1.4 Criteria Guidelines

Aggregate Validation Checks

In the Peer Exchange on Travel Model Validation Practices held in Washington, D.C. on May 9, 2008, a general consensus of participants was:

> There was some agreement that setting validation standards for matching traffic counts, transit boardings, and screenline crossings can be a double-edged sword. While standards can be used to help determine relative model accuracy, they also can encourage over-manipulation to meet the standards. This can be especially true if project rankings or construction funds are based on absolute values rather than relative results. While almost any travel model can be manipulated to attain a specified validation standard, it is important to emphasize the use of appropriate methods to meet the standard. Methods used to achieve a reasonable match between modeled and observed traffic volumes can be as important as the reasonableness of the match itself. Therefore, model validation should focus on the acceptability of modeling practices in addition to attaining specified standards. A model validation that matches specified trip assignment standards within a reasonable range using valid modeling procedures is better that a model that

matches observed volumes with a tighter tolerance using questionable modeling procedures.[34]

Based on the above, this chapter reports some guidelines that have been used by various states and agencies. Specifically, Table 9.2 lists some example guidelines used for the match between modeled and observed VMT for Ohio and Florida. Figure 9.8 summarizes %RMSE guidelines used in Ohio, Florida, and Oregon. The Michigan Department of Transportation (MDOT) has targets of 5 percent and 10 percent for screenlines and cutlines, respectively, for percent differences in observed and estimated volumes by screenline. Figure 9.9 shows the maximum desirable deviation in total screenline volumes according to the observed screenline volume originally cited in *Calibration and Adjustment of System Planning Models*, produced by the FHWA in December 1990, and referenced in a number of documents, including the NCHRP Report 255, and the 1997 version of this manual. **The guidelines in this section should *not* be construed as standards; matching or exceeding the guidelines is *not* sufficient to determine the validity of a model.**

Disaggregate Validation Checks

There are no specific criteria guidelines associated with disaggregate traffic assignment checks described above.

[34] *Peer Exchange on Model Validation Practices White Paper*, prepared by Cambridge Systematics, Inc., for the Federal Highway Administration, December 18, 2008, page 5, http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/tmip/peer_exchange/2008-05-09/model_validation.pdf, accessed September 28, 2009.

Table 9.2 Example VMT Guidelines by Functional Class and Area Type

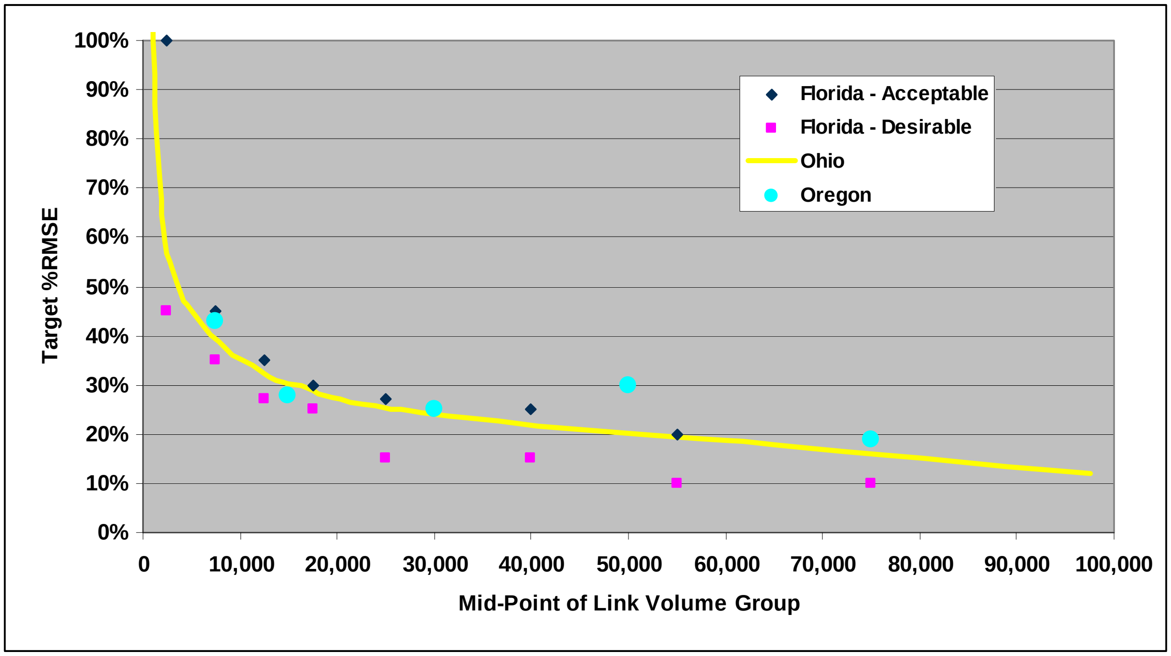
	Modeled Versus Observed VMT				
Stratification	**Ohio**[a]	**Florida**[b]		**Michigan**[c]	**FHWA-1990**[c]
Functional Class		**Acceptable**	**Preferable**		
Freeways/Expressways	±7%	±7%	±6%	±6%	±7%
Principal Arterials	±10%	±15%	±10%	±7%	±10%
Minor Arterials	±10%	±15%	±10%	±10%	±15%
Collectors	±15%	±25%	±20%	±20%	±20%
All Links		±5%	±2%		
Area Type					
CBD	±10%	±25%	±15%		
Fringe	±10%	±25%	±15%		
Urban	±10%	±25%	±15%		
Suburban	±10%	±25%	±15%		
Rural	±10%	±25%	±15%		

[a] Giaimo, Gregory, Travel Demand Forecasting Manual 1 – Traffic Assignment Procedures, Ohio Department of Transportation, Division of Planning, Office of Technical Services, August 2001.

[b] FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards: Model Validation Guidelines and Standards, prepared by Cambridge Systematics, Inc., for the Florida Department of Transportation Systems Planning Office, December 31, 2007, Table 3.9, page 3-16.

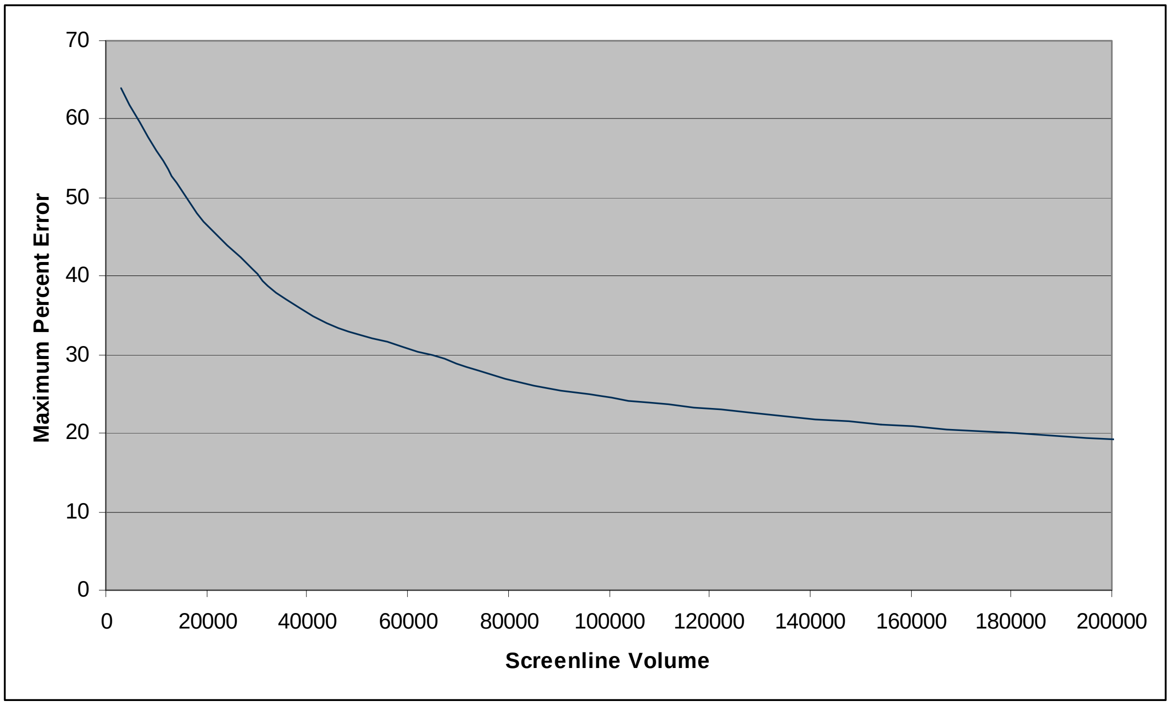
[c] The FHWA Travel Model Improvement Program Workshop over the Web, The Travel Model Development Series: Part I–Travel Model Estimation, prepared by Cambridge Systematics, Inc., June 9, 2009, Slide 11, http://tmip.fhwa.dot.gov/sites/default/files/presentation_8_with_notes.pdf, accessed November 29, 2009.

Figure 9.8 Example %RMSE Guidelines



Sources: Ohio: Giaimo, Gregory, Travel Demand Forecasting Manual 1–Traffic Assignment Procedures; Florida: FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards: Model Validation Guidelines and Standards; and Oregon: FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards, Draft Technical Memorandum 1.

Figure 9.9 Example Maximum Desirable Deviation in Total Screenline Volumes Guidelines



Source: Calibration and Adjustment of System Planning Models, FHWA, December 1990.

9.1.5 Reasonableness and Sensitivity Testing

Reasonable ranges of VMT per household are 40 to 60 miles per day for large urban areas and 30 to 40 miles per day for small urban areas. The 1990 NPTS reported an average of 41.37 vehicle miles traveled per household daily. The average increased to 58.05 vehicle miles of travel in the 2001 NHTS (although differences in the survey methods account for some of the increase). Reasonable ranges of VMT per person are 17 to 24 miles per day for large urban areas and 10 to 16 miles per day for small urban areas.

Traffic assignment techniques vary from region to region. Based on a review of the model documentation of assignment procedures used by 40 different MPOs throughout the country:

- About 70 percent use time-of-day traffic assignment procedures;
- 75 to 80 percent perform class-based assignment techniques; and
- 20 to 30 percent perform speed equilibration for some of the assigned time periods.

Table 9.3 summarizes the ranges of coefficients and exponents of BPR-like volume delay functions as reported by 18 of the MPOs. The BPR-like function estimates the congested travel time on a link using the following formula:

$$Time_{final} = Time_{initial} \times \left[1.0 + \alpha\left(\frac{V}{C}\right)^{\beta}\right]$$

Where:

$Time_{final}$ is the final, congested travel time on a link;

$Time_{initial}$ is the initial, or starting, travel time on a link;

V is the assigned volume on a link;

C is the capacity of the link (at level of service E); and

α and β are model coefficients

Sensitivity testing of traffic assignment procedures can be performed by making changes to the networks or input trip tables used for assignment. Several approaches are as follows:

Table 9.3 Range of Reported BPR-Like Assignment Parameters (18 MPOs)

	α		β	
Facility Type	**Minimum**	**Maximum**	**Minimum**	**Maximum**
Freeways	0.10	1.20	1.90	10.00
Arterials	0.15	1.00	2.10	4.00

- **Regional sensitivity –** Check reasonableness in change in VMT to changes in total trips. Increase (factor) trips by a factor (e.g., 1.5) and check to see that total VMT changes by a similar factor. If there is little congestion in the region, VMT should increase by a similar factor. If there is substantial congestion, VMT should increase by more than the factor.

- **Localized sensitivity –** Modify key network elements and review assignment results for changes and reaction to network elements (using a fixed trip table). For example, remove a key bridge or limited access facility and review the impact on traffic using volume difference plots between the original and modified alternatives.

- **Over-sensitivity –** For congested networks, make a minor change to a network (e.g., add a lane of traffic to a minor arterial link) and reassign a fixed trip table using same number of iterations and closure criteria. Review the impact on traffic using volume difference plots between the original and modified alternatives. Traffic impacts should be very localized.

9.1.6 Troubleshooting Strategies

Since traffic assignment is the culmination of the modeling process, issues can easily be related to previous steps in the modeling process. It is, however, always valid to start the troubleshooting with the traffic assignment step and work backwards through the modeling process. Table 9.4 provides some troubleshooting strategies for common issues that might occur with a traffic assignment.

Table 9.4 Troubleshooting Strategies for Issues with Traffic Assignment

Issue	Potential Troubleshooting Strategies
1. Low, high, or unrealistic base year modeled link volumes compared to traffic counts	• Check network coding (speeds, capacities, etc.) on these links, nearby/adjacent links, and links on competing paths • Check TAZ connections and loading at centroids • Check traffic count data
2. Uneven facility loading on parallel competing routes	• Review centroid connections • Review facility and area type coding and input starting speeds for assignments • Review zone structure and number of zones – may need to have finer spatial resolution • Review final congested speeds and volume-delay functions
3. Travel times not representative of observed data	• Review facility and area type coding and input starting speeds for assignments • Review final congested speeds and volume-delay functions
4. Links with zero assigned volume	• Check network coding (including nearby or competing links) for continuity, stub links, centroid connector locations, and attributes such as free-flow speeds and capacities
5. Links with very high assigned volume/ capacity ratios	• Check network coding (including nearby or competing links) for centroid connector locations and attributes such as free-flow speeds and capacities

9.1.7 Forecast Checks

The forecast year validation checks for traffic assignment should concentrate on comparisons of the forecast year model results to the base year model results. The base year observed data are no longer directly considered. Unlike the base year comparisons, however, the objective is not to achieve a close match between the forecast and base year results, but rather to ensure that the differences and trends are reasonable. For example, it may be reasonable to expect that VMT per capita increases somewhat over time especially in congested regions due to increased circuitry of travel.

The main comparisons are similar to the comparisons previously done between base year model results and observed data. These may include regional, subregional, and corridor specific checks. Examples of regional and subregional checks include:

- VMT per capita;
- Total VMT by functional class;
- Average congested speeds by functional class;
- Changes in VMT by functional class; and
- Changes in volumes crossing screenlines, cutlines, and cordon lines.

Examples of corridor-level checks include:

- Difference plots of future versus base year traffic; and
- Comparisons of speeds on facilities.

Traffic for specific facilities should not always be expected to increase. Facilities that are congested in the base year may not be able to handle significantly more traffic in the future or capacity improvements or new roadways in other areas might minimize increases in traffic on specific facilities.

9.2 Transit Assignment Checks

Traditional transit assignment procedures have focused on the assignment of peak and off-peak period trips in production-attraction format in an effort to reproduce daily transit boardings by line and, in many cases, the ridership at maximum load points along the line. For regional travel forecasts, the output of transit assignments may be somewhat less rigorous than the requirements for traffic assignments. The regional assignments have been used to determine information such as the number of transit vehicles required based on the frequency of service required to serve the forecast transit demand.

The amount of scrutiny received by transit assignments may increase substantially when a region applies for FTA Section 5309 New Starts funds. The FTA encourages rigorous checking of transit networks, transit path-building procedures, and transit assignment results.

Generally available transit assignment procedures include all-or-nothing, all shortest paths, and a number of multipath assignment procedures. The multipath transit assignment procedures are heuristic procedures used to represent the optional path choices and path use in robust transit systems rather than transit path-switching due to capacity constraints. Capacity constrained transit assignment techniques are rarely required. Few regions reach crowding to an extent that people switch from one transit path or mode to another to avoid the over-crowding. In cases where this does occur, ad hoc techniques such as "shadow pricing at park and ride lots" are used to "move" transit ridership to different lines.

Some regions developing activity-based travel models have moved from peak and off-peak transit assignments in production-attraction format to true time-of-day assignments in origin-destination format. This change can be considered to be evolutionary; not revolutionary. The transit assignment validation checks for origin-destination-based transit assignments are similar to those used for more tradition transit assignments in production-attraction format.

As discussed in Chapter 7, transit assignment validation is closely related to mode choice model validation, as it regards transit mode choices. Issues identified during checks of transit assignment results may be caused by issues with the mode choice model, and vice versa, and issues with both model components may be related to transit path building and network skimming procedures.

9.2.1 Sources of Data

The primary source of data for transit assignment validation is the transit operator. The most generally available data are count data such as boardings by line and park-and-ride lot utilization counts. Some regions will also have on-board survey data available for validation.

Boarding Count Data

Most transit operators collect boarding count data by line on a continuous basis through the use of recording fare boxes or the performance of periodic counts. In some cases, the data may be available by time-of-day. Route-level boarding count data can be easily aggregated by mode or by corridor.

Some transit operators are installing Automated Passenger Counters (APCs) on their transit vehicles or perform periodic "boarding and alighting" counts on transit lines. If both the numbers of boardings and the numbers of alightings by transit stop are available, route profiles can be constructed for the lines. Detailed boarding and alighting data at bus stops also provide the means for developing route profiles; screenline, cutline, and cordon line counts; and estimates of passenger-miles of travel (PMT).

Park-and-Ride Lot Utilization

Regions that have an established park-and-ride system may collect parking lot utilization data for the various lots. The data collected may range from number of spaces used on a daily basis to the number of vehicles parking at the lot on a daily basis to license plate surveys of parking lots. Vehicle counts at park-and-ride lots are superior to counts of used parking spaces since the vehicle counts provide a clearer picture of park-and-ride lot demand.

Transit Rider Survey Data

Transit rider survey data, often collected as on-board survey data, provide a wealth of information for detailed transit assignment validation including transfer rates, numbers of linked trips, and access and egress modes. Survey data that represent all transit service in the modeled region provide the information necessary to develop "observed" transit trip tables. The development of the observed trip tables requires careful expansion of the on-board survey data to match boarding counts.

When observed trip tables are available, it is possible to focus the validation on the actual transit assignment procedures since the validation will not be impacted by the veracity of the trip tables produced by the rest of the modeling process. This is in contrast to the traffic assignment validation process where it is not possible to collect sufficient data to develop observed auto trip tables for a general validation of the traffic assignment process.

Transit operators who have received FTA New Starts funding are required to perform on-board surveys before and after the construction to determine who is using and benefiting from the new system. The availability of before and after trip tables provides unique data for transit assignment validation, including the ability to perform temporal validations.

Other Data

In some areas, operators may have other useful data available, especially where automated passenger counting or fare collection is performed. For example, transit systems that use "smart cards" or similar technology may have information on boarding and alighting stations of passengers, which could be compared to model results.

9.2.2 Aggregate Checks

Aggregate data checks may be performed using trip tables resulting from the modeling process through mode choice and, possibly, time-of-day modeling, or using transit trip tables from comprehensive on-board survey. If observed trip tables are available, tests should be performed using those tables since differences between modeled and observed transit validation measures can be more fully attributed to the transit networks and transit assignment process.

Boarding Count Checks

Most aggregate transit assignment checks begin with the comparison of modeled to observed transit boardings. In addition to total system boardings, these comparisons may include boardings by line and by mode. The checks may be performed by time-of-day. Validation checks typically consist of comparing absolute and relative differences between modeled and observed boardings by line. Since most regions have relatively few transit lines, checks by line are typically reported for each line. The reports may be stratified by percent difference to facilitate diagnosis of transit assignment problems.

Comparison of modeled to observed boardings at major transfer points provides another set of validation checks. The major transfer points may include park-and-ride lots, fixed guideway transit stations (e.g., light-rail stations), and bus transit centers or "pulse-points."

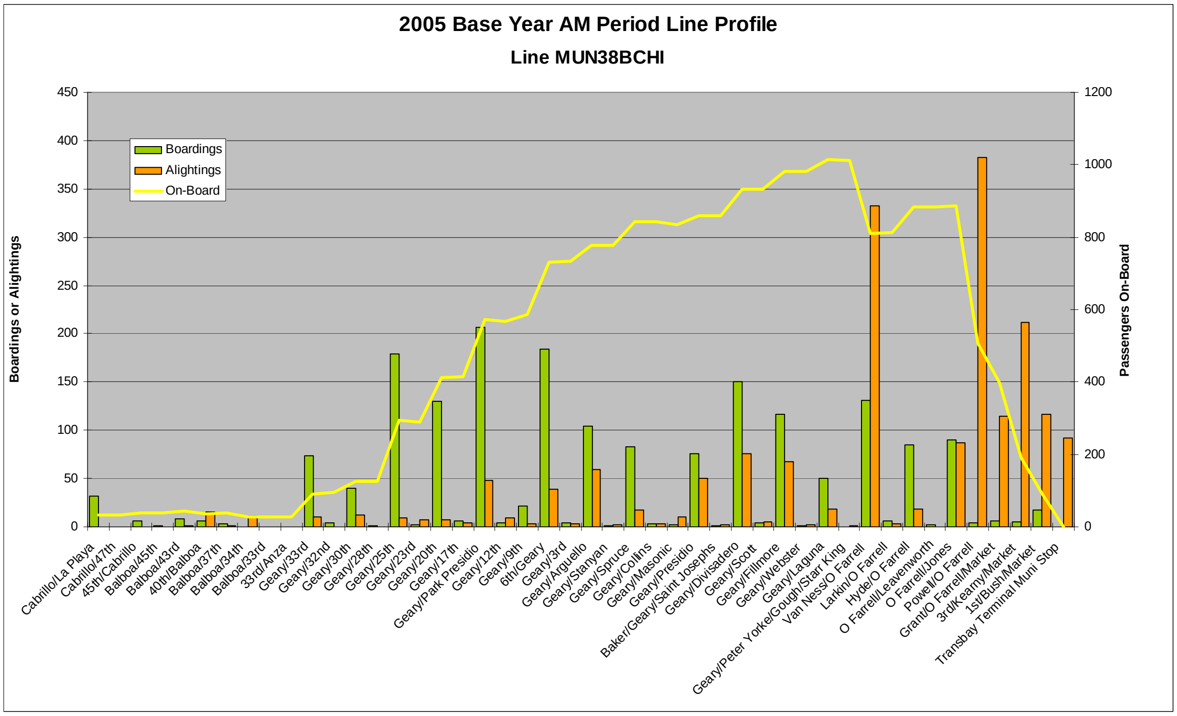
The assignment of an "observed" transit trip table (based on expanded data from a transit rider survey) can be valuable in providing an "in-between" data point for transit assignment validation. If the modeled boardings resulting from the assignment of the "observed" transit trip table match the observed boardings reasonably well, but the modeled boardings resulting from the assignment of the transit trip table from the mode choice model do not match up well with the observed boardings, issues with the mode choice model (or preceding models such as trip distribution) may be indicated. If the results from assignments using both trip tables ("observed" and from the mode choice model) match each other well but not the observed boardings, there may be issues with the transit network or path building procedures (although checks of the observed data, boardings and transit survey, should also be performed).

Boarding- and Alighting-Based Checks

If detailed boarding and alighting data are available, it is possible to construct observed transit route profiles such as the example shown in Figure 9.10. This information provides the means to compare modeled to observed volumes along transit lines. Modeled line profiles may be compared to observed profiles for selected lines.

Line profiles can be used to determine observed transit riders on transit screenlines, cutlines, and cordon lines. Comparison of modeled transit volumes on screenlines, cutlines, and cordon lines resulting from the assignment of modeled trip tables to observed volumes is useful in determining the quality of the trip distribution and mode choice procedures. If transit volumes resulting from the assignment of observed trip tables are compared to the observed volumes, the results might demonstrate improper diversions to alternative transit paths in robust transit systems. Such results might lead to modifications of the transit path-building and assignment parameters or the transit network coding.

Figure 9.10 Example Route Profile



Modeled PMT for the region, by line, by mode, by access mode, or by time-of-day can be compared to observed PMT when detailed boarding and alighting counts are available.

Transit Rider Survey-Based Checks

If a transit rider survey is available, the regional transfer rate or boardings per linked trip can be estimated. This information can also be estimated from boarding counts provided the operator provides transfers and records boardings by fare payment type. Modeled boardings per linked trip can be estimated from the transit assignment results. As with previous aggregate checks, this comparison can be made based on the assignment of observed transit trip tables or based on the assignment of modeled trip tables.

9.2.3 Disaggregate Checks

The following checks must be performed using data collected in a comprehensive transit rider survey. The following checks are not truly disaggregate as defined for discrete choice models, but are substantially more detailed than the aggregate checks described above. In effect, these checks involve comparisons of transit paths reported by travelers in the survey to modeled paths. The disaggregate checks are based on the analysis of individually reported transit trips rather than the assignment of an observed transit trip table for the region.

The reported trips should be compared to transit paths build using procedures consistent with the transit assignment process. For example, if transit trips using walk access are forecasted by the mode choice model and assigned separately for local bus and rail, individually reported trips for travelers using walk access to local bus only should be compared to the modeled walk access to local bus transit path information. Likewise, the individually reported trips for travelers using rail in their transit trip should be compared to the modeled walk access to rail transit path information. Conversely, if the mode choice model forecasts only total walk access trips, the individually reported trips for all travelers using walk access should be compared to the modeled walk access to transit path information. In this case, it might be worthwhile to check the prediction success of boardings by mode (e.g., local bus and rail for this example) rather than total boardings on the interchanges.

Comparison of modeled to reported transit paths can be used to prepare prediction success tables of the transit path-builder and path-building parameters used for the assignment process. While modeled paths could be compared to reported paths and the results summarized in "pass-fail" form, such an approach could be extremely time consuming. The process can be automated to summarize key variables. Table 9.5 is an example of a prediction success table for modeled to reported boardings on individual transit paths and Table 9.6 shows a summary of the results.

Table 9.5 Example Prediction Success Table for Transit Assignment

		Skimmed Boardings for Reported Interchanges			
		No Path	**1**	**2**	**3+**
On-Board Survey Reported Boardings	**No Path**	0	0	0	0
	1	47	640	230	1
	2	6	217	102	18
	3+	0	26	8	50

Table 9.6 Example Prediction Success Table Summary for Transit Assignment

		Percent with Skimmed Boardings:		
Assignment	**Number of Linked Trips**	**Equal Reported Boardings**	**Greater Than Reported Boardings**	**Less Than Reported Boardings**
Walk Access	854	67%	23%	9%
Drive Access	424	67%	25%	7%
All Trips	**1,278**	**67%**	**24%**	**9%**

9.2.4 Criteria Guidelines

Aggregate Validation Checks

The same caveat regarding setting guidelines for aggregate traffic assignment validation checks can be made for aggregate transit validation checks. Setting guidelines is double-edged sword that may lead to over-manipulation of transit assignment procedures. Consequently, this chapter reports some guidelines that have been used by various states and agencies. **The guidelines in this section should *not* be construed as standards; matching or exceeding the guidelines is *not* sufficient to determine the validity of a model.**

It should be noted that the FTA does not specify guidelines for the New Starts program other than that the overall modeling process should "tell a coherent story." The FTA focus is on reasonable reproduction of the transit network and transit travel times and reasonableness of predicted changes between current and future ridership coupled with reasonableness of changes between future base and future build alternatives.

What is being validated must be considered. If observed trip tables from a comprehensive transit rider survey are being assigned and used as a basis for the validation, much more emphasis is being placed on transit assignment procedures (although there is some consideration of the veracity of the "observed" trips tables and expansion factors). In this case, a "tight" validation might be desired. Alternatively, if a modeled trip tables from mode choice and transit time-of-day processing is being assigned to provide the modeled transit boardings and transit flows for validation, the validation actually covers the entire modeling process to that point in addition to the validation of the transit assignment process. In this case, desired criteria might be less stringent.

Example transit assignment validation results for several areas are shown in Tables 9.7 through 9.10. Tables 9.9 and 9.10 show example transit screenline validation results and guidelines.

PMT for transit assignment is analogous to VMT for traffic assignment. As a result, any regional VMT guideline set for traffic assignment results might be used for regional modeled PMT to observed PMT. For example, if regional guidelines suggest that regional VMT be within ±5 percent of the observed VMT, the same guideline might be considered for the transit assignment.

Disaggregate Validation Checks

There are no specific criteria guidelines associated with disaggregate transit assignment checks described above.

Table 9.7 Example Transit Validation Results for Sacramento Region

Type of Trip or Boarding	1999 Observed[a]			SACMET01 2000			Validation Ratios		
	Peak	Off-Peak	Total	Peak	Off-Peak	Total	Peak	Off-Peak	Total
Transit Linked Trips									
Walk (RT Only)	26,258	22,890	49,148	20,975	26,628	47,603	0.8	1.16	0.97
Drive (RT Only)	8,738	2,331	11,069	10,937	3,154	14,091	1.25	1.35	1.27
RT Subtotal	34,996	25,221	60,217	31,912	29,782	61,694	0.91	1.18	1.02
Other Bus[b,c]	N/A	n/a	5,583	n/a	n/a	8,384	n/a	n/a	1.5
Total	**N/A**	**N/A**	**65,800**	**N/A**	**N/A**	**70,078**	**N/A**	**N/A**	**1.07**
Transit Boardings by Bus/LRT									
LRT	15,244	11,673	26,917	14,583	9,754	24,337	0.96	0.84	0.9
RT Bus	33,081	25,442	58,523	28,547	30,296	58,843	0.86	1.19	1.01
RT Subtotal	48,325	37,115	85,440	43,130	40,050	83,180	0.89	1.08	0.97
Other Bus[b]	n/a	n/a	6,978	4,791	4,789	9,580	n/a	n/a	1.37
Total	**N/A**	**N/A**	**92,418**	**47,921**	**44,839**	**92,760**	**N/A**	**N/A**	**1**
LRT Boardings (By Access Mode at Production End of Trip)									
Transfer	4,332	4,404	8,736	3,188	3,164	6,352	0.74	0.72	0.73
Walk	4,006	5,394	9,400	2,412	3,838	6,250	0.6	0.71	0.66
Drive	6,905	1,876	8,781	8,983	2,752	11,735	1.3	1.47	1.34
Total LRT Boardings	**15,243**	**11,674**	**26,917**	**14,583**	**9,754**	**24,337**	**0.96**	**0.84**	**0.9**
Bus Boardings (By Access Mode at Production End of Trip)									
Transfer (RT Only)	8,997	7,491	16,488	n/a	n/a	12,146	n/a	n/a	0.74
Walk (RT Only)	22,252	17,496	39,748	n/a	n/a	44,012	n/a	n/a	1.11
Drive (RT Only)	1,833	455	2,288	n/a	n/a	2,685	n/a	n/a	1.17
RT Subtotal	24,085	17,951	58,524	n/a	n/a	58,843	n/a	n/a	1.01
Other Bus[b]	n/a	n/a	6,978	n/a	n/a	9,580	n/a	n/a	1.37
Total Bus Boardings	**N/A**	**N/A**	**65,502**	**N/A**	**N/A**	**68,423**	**N/A**	**N/A**	**1.04**

Sources: DKS Associates, 2002; and *Sacramento Regional Travel Demand Model Version 2001 (SACMET 01)*, prepared by DKS Associates for Sacramento Association of Governments, prepared by DKS Associates, March 8, 2002, Table 43.

[a] RT numbers based on 1999 On-Board Surveys conducted by RT and SACOG for RT system only. Other bus operators based on total boardings estimates provided by operators.

[b] Includes Yolobus, Roseville Transit, Placer County Transit, Folsom Transit, El Dorado Transit. Excludes free or near-free operators (Unitrans, CSUS Shuttle)

[c] Trips for other bus computed from boardings, using data from 1994 On-Board Transit Survey.

Table 9.8 Example Transit Validation Results for Seattle Region

	2000 Modeled Boardings			2000 Observed Boardings	
Transit Operator	**AM**	**MD**	**Daily**	**Daily**	**Percent Difference**
King County Metro	92,940	77,627	294,226	329,913	-11%
Pierce Transit	9,987	11,440	36,661	45,265	-19%
Community Transit and Everett Transit	10,070	7,662	30,660	33,318	-8%
Kitsap Transit	4,403	3,967	14,410	11,889	21%
Washington State Ferries	11,372	2,114	23,979	21,000	14%
Sound Transit	10,006	8,900	32,560	see note	n/a
Total	**138,778**	**111,710**	**432,497**	**441,385**	**-2%**

Source: *PSRC Travel Model Documentation (for Version 1.0) – Updated for Congestion Relief Analysis*, prepared by Cambridge Systematics, Inc., for Washington State Department of Transportation and Puget Sound Regional Council, September 2007.

Notes: Observed boardings are from the National Transit Database (NTD). Sound Transit boardings were reported in NTD under other operators, King County Metro, Pierce Transit, and Community Transit.

Table 9.9 Example Transit Assignment Validation Guideline for State of Florida

	Benchmarks	
Validation Statistic	**Acceptable**	**Preferable**
Regional Estimated-over-Observed Transit Trips (Boardings)	+/- 9%	+/- 3%
Transit Screenlines	+/-20%	+/-10%
Transit Line Ridership: <1,000 Passengers/Day	+/-150%	+/- 100%
Transit Line Ridership: 1,000-2,000 Passengers/Day	+/- 100%	+/- 65%
Transit Line Ridership: 2,000-5,000 Passengers/Day	+/- 65%	+/- 35%
Transit Line Ridership: 5,000-10,000 Passengers/Day	+/- 35%	+/- 25%
Transit Line Ridership: 10,000-20,000 Passengers/Day	+/- 25%	+/- 20%
Transit Line Ridership: >20,000 Passengers/Day	+/- 20%	+/- 15%

Source: *FSUTMS-Cube Framework Phase II – Model Calibration and Validation Standards: Model Validation Guidelines and Standards*, prepared by Cambridge Systematics, Inc., for Florida Department of Transportation Systems Planning Office, December 31, 2007.

Table 9.10 Example Transit Screenline Results for Seattle Region

Screenline Location	1999 Observed	2000 Modeled	Difference	Percent Difference
132nd SW, Snohomish County	5,825	6,883	1,058	18%
Snohomish County Line West	10,590	11,449	859	8%
Snohomish County Line East	2,010	1,582	-428	-21%
Ship Canal Bridges	65,970	56,160	-9,810	-15%
Lake Washington Bridges	20,670	21,999	1,329	6%
Newport Eastside	3,430	4,948	1,518	44%
South Spokane Street	60,100	32,347	-27,753	-46%
West Seattle Bridges	21,500	20,752	-748	-3%
South 188th Street, King County	21,170	10,703	-10,467	-49%
Pierce County Line	6,860	4,780	-2,080	-30%
40th Street, Tacoma	9,300	2,544	-6,756	-73%
Eastside, North of I-90	9,850	3,916	-5,934	-60%
Eastside, East of I-405 (E-W	2,760	2,258	-502	-18%
Eastside, North of Kirkland	8,100	6,602	-1,498	-18%
Eastside, North of Renton	2,630	3,209	579	22%
South King County (E-W Movements)	10,260	3,433	-6,827	-67%
Subtotals				
King County – Seattle	199,670	145,394	-54,276	-27%
King County – Eastside	26,770	20,932	-5,838	-22%
Pierce County	16,160	7,324	-8,836	-55%
Snohomish County	18,425	19,914	1,489	8%
All Screenlines	**261,025**	**193,564**	**-67,461**	**-26%**

Source: *PSRC Travel Model Documentation (for Version 1.0) – Updated for Congestion Relief Analysis*, prepared by Cambridge Systematics, Inc., for Washington State Department of Transportation and Puget Sound Regional Council, September 2007.

9.2.5 Reasonableness and Sensitivity Testing

Perhaps the best reasonableness test that can be applied to transit assignment results is the application of the "tell a coherent story" philosophy to the transit assignment. In effect, the transit assignment process should "tell a coherent story" regarding how transit riders behave. Beyond that suggestion, there are several reasonableness checks that can be made:

- Are the transit path-building parameters used for the transit assignment consistent with the mode choice model coefficients?

- Does the number of boardings per linked trip (or transfer rate) make sense? Boardings per linked trip are typically in the range of 1.2 to 1.6 with the higher rates in regions with grid-based bus systems and fixed guideway transit modes (e.g., light rail, heavy rail, or bus rapid transit).

- Do maximum load point locations make sense (even if observed locations for maximum load points are not available)? For example, maximum load points for radial transit lines focused on a central business district or some other major generator should be reasonably near the major generator. For cross-town routes, the maximum load point should probably be closer to the central portion of the route.

Sensitivity testing of transit assignment procedures can be performed by making changes to the networks or input trip tables used for assignment. Some approaches include:

- **Regional sensitivity –** Check the reasonableness in changes in total boardings to changes in total trips. Increase trips by a factor (e.g., 1.5) and check to see that total boardings change by a similar factor.

- **Localized sensitivity –** Modify speeds or headways on selected routes and observe the changes in boardings (especially in areas where there is "competition" among transit routes). Do faster or more frequent routes attract more riders? Remove routes and observe change in ridership on other routes.

- **Mode sensitivity –** If walk to rail (or walk to premium transit) is assigned separately from walk to local bus, increase rail trips on specific interchanges that must use background bus to access rail by a known number of linked trips. Verify that rail boardings increase by at least (or exactly) the increase in the number of linked trips.

9.2.6 Troubleshooting Strategies

Transit assignment, like traffic assignment, is the culmination of the modeling process. As a result, issues can easily be related to previous steps in the modeling process. However, unlike traffic assignment, it might be possible to isolate transit assignment issues to the transit assignment process if an observed transit trip table from an on-board survey is available. Table 9.11 provides some troubleshooting strategies for common issues that might occur with a traffic assignment. Also refer to Table 7.6, which presents the analogous strategies for the mode choice model.

Table 9.11 Troubleshooting Strategies for Issues with Transit Assignment

Issue	Potential Troubleshooting Strategies
1. Low or high boardings/ridership compared to route/stop boardings	• Check network coding (stops, etc.) on the affected routes/stops, nearby/adjacent routes, and competing routes • Check transit access links • Check run times, speeds, and/or dwell times for routes • Check level of zonal resolution and transit walk access percentages • Check trip tables for consistency between trips in corridor and observed boardings • Modify path-building/assignment parameters • If using multi-path assignment procedures, investigate changes in route "combination" factors • Investigate changes to transfer penalties • Investigate changes to relationships between wait time, out-of-vehicle time, in-vehicle time, and transit cost
2. Low or high boardings per linked trip	• Review walk network assumptions • Investigate changes to transfer penalties • Modify assignment procedures • Increase market segmentation • Modify path-building/assignment parameters • If using multi-path assignment procedures, investigate changes in route "combination" factors • Investigate changes to transfer penalties • Investigate changes to relationships between wait time, out-of-vehicle time, in-vehicle time, and transit cost

9.2.7 Forecast Checks

Certain basic statistics such as the number of boardings per linked trip and PMT per linked trip should remain relatively constant between the base year and a future year unless, of course, the transit system has been modified in such a way that it directly impacts one of those statistics. For example, the introduction of light-rail transit in a region would probably increase the boardings per linked trip and the introduction of commuter rail might increase the PMT per linked trip.

The FTA has suggested a number of checks that should be used when producing ridership forecasts for a Section 5309 New Starts analysis, but these suggestions would be applicable for any future transit assignment (regardless of whether it is for a New Starts project). Figures 9.11a and 9.11b summarize the FTA suggestions for forecast checks.

Figure 9.11 FTA New Starts-Based Forecasting Checks

Demonstrating Reasonable Predictions of Change

- Models should provide reasonable predictions of change
  - ✓ Between today and a future no-build condition
  - ✓ Between a future no-build condition and a realistic alternative (i.e., a change in the transportation system)
- To be useful, tests of reaction to change must be done through applications of the model in full production mode
  - ✓ Findings can highlight problems not prevalent in base year conditions

(a)

Common Tests for Reasonable Forecasts

No.	Compare model results from…	…to the results from…	This compares the…	The results are different because of …
1	Previously validated year	Base year validation	Past to the present	Changes in demographics and employment and transportation supply
2	Base year validation	Future year no-build	Present to the future	Demographic and employment forecasts
3	Future year no-build	Future year TSM	The future to a modestly-changed future	Transportation supply (modest)
4	Future year TSM	Future year Build	The modestly-changed future to a future with a big project	Transportation supply (major)

(b)

Source: *Shining a Light Inside the Black Box (2): Model Testing*, Travel Model Improvement Program Webinar, March 11, 2008, presentation slides 33 and 34, http://tmip.fhwa.dot.gov/sites/tmip.fhwa.dot.gov/files/presentation_bb2.pdf, accessed October 5, 2009.

10.0 Temporal Validation and Sensitivity Testing

10.0 Temporal Validation and Sensitivity Testing

Temporal validations of travel models have been performed by many agencies as a matter of practice. Quite often, an existing travel model is applied for an updated base year to show that the model is still "valid." For example, a regional model may have been most recently estimated using data from a household travel survey collected in 1997. The model may have been validated against independent data collected for that purpose for the original base year, 1997. The agency maintaining the model might then "validate" the model against traffic and transit ridership data collected, say, every five years until a new household survey is collected. Thus, model system "validations" might exist for 2002 and 2007 with the next major data collection and model estimation scheduled for 2012. The 2002 and 2007 validations might include adjustments to the models to better reproduce observed traffic and transit ridership.

Temporal validations such as the aforementioned are important for systemwide model validations, particularly if sufficient time or transportation system changes exist between the years selected for the validation. They may be necessary for maintaining credibility of the travel models with local officials or may be required by an MPO or state DOT. The periodic systemwide validations are often crucial for demonstrating the need to collect new travel survey data and estimate updated travel models.

While these periodic checks to ensure that established travel models continue to reasonably reproduce observed traffic and transit ridership are important, they represent a special case of the temporal validation and sensitivity tests described in this chapter. The primary focus of this chapter is an expanded concept of temporal validation and sensitivity testing that can be performed when travel models are estimated and calibrated. The ideas described in this chapter are generally applicable for both trip-based and activity-based travel models.

Several definitions guide the concepts described in this chapter:

- *Forecast* is any application of the travel model for any year after the model calibration year;
- *Backcast* is any application of the model for a year prior to the calibration year;
- *Temporal validation tests* are comparisons of model forecasts or backcasts against observed travel data; and

- *Sensitivity tests* are evaluations of model forecasts for years or alternatives for which observed data do not exist.

Thus, if a model was calibrated using 1997 data, applications of the model for 2002 and 2007 would represent forecasts even though those years are now history. Comparisons of the forecasts for 2002 or 2007 against observed data such as traffic counts would be temporal validation tests. Likewise, if the model had been calibrated using 2007 data, applications of that model using the 1997 and 2002 data would represent backcasts, and comparisons to observed 1997 or 2002 data would also be temporal validation tests.

10.1 Temporal Validation

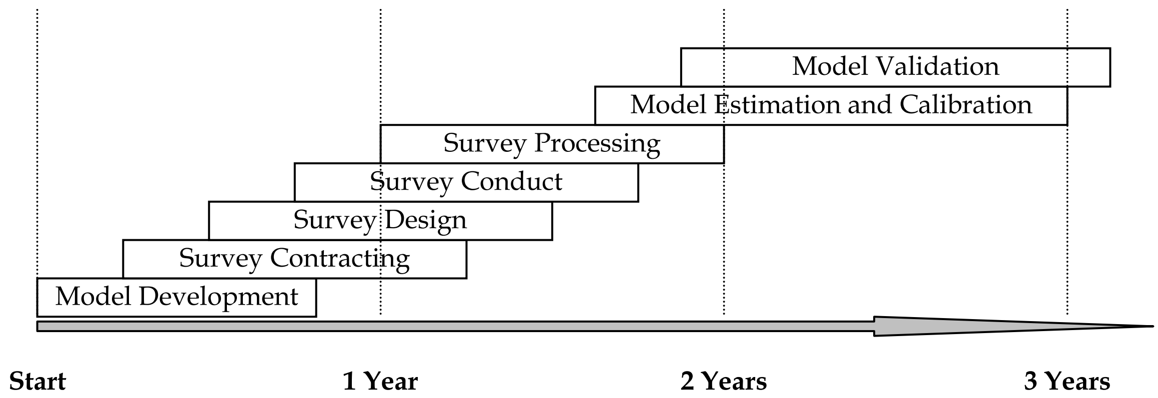
10.1.1 Timeline for Calibration and Temporal Validation

Temporal validations should be performed when travel models are developed. Figure 10.1 shows an example timeline for a model development process including the collection of travel survey data. As shown in Figure 10.1, the time from the initiation of model development to model validation can easily be three years. As recommended in Chapter 2, the development of a validation plan should take place in conjunction with the development of a model development plan. Based on the model validation plan, validation data would ideally be collected in parallel with the survey. However, with the short timeframe for data collection and model development shown in Figure 10.1, a backcast would be the most likely procedure for temporal validation.

There are, of course, many variations to the example shown in Figure 10.1 that may compress or expand the timeline. More complicated data collection schemes, resource constraints, or extension of the model estimation and calibration might extend the timeline to, say, five or more years. In such a case, it might be possible to assemble some validation data toward the end of the timeline that would allow a temporal validation using a forecast to, say, the fifth or sixth year after the project initiation.

A variation of the backcast/forecast option for model validation has been used by the Denver Regional Council of Governments (DRCOG). DRCOG performed several travel surveys between 1997 and 1999 and refreshed the various components of their trip-based travel models between 2000 and 2004. Development of an activity-based model using the 1997 household survey data was initiated in 2006. Since almost 10 years had elapsed between the initial data collection and the activity-based model development, DRCOG estimated the models using the 1997 data, calibrated the models to match 2005 conditions, and backcast to 1997 for temporal validation.

Figure 10.1 Example Timeline for Model Development



10.1.2 Temporal Validation Data

Temporal validation tests are, of course, dependent upon data availability. Many temporal validations focus on matching traffic counts and transit boardings due to the general availability of those data. The validation tests are limited to those used for trip assignment validation or overall model system validation. While such validations are important and should be performed, they do not validate the various model components over time. Validating other model components require additional data to supplement traffic count and transit boarding data.

Sources of Backcast Validation Data

Several primary sources of backcast data, other than historical traffic counts and transit boarding counts, may exist for a region. The sources include previous travel survey data, census data, historical travel model applications (especially from previous validations), and written reports.

Previous Travel Survey Data

If data from a previous travel survey are well documented and archived, many of the model validation tests described in this manual may be possible when models are updated (based on a new data collection effort). If network and socioeconomic data used for model development in conjunction with the previous travel survey are also archived, it might be possible to consider a full range of temporal model validation tests.

Since modeling procedures and model applications software change over time, the existence of archived networks and socioeconomic data does not guarantee that a backcast can be performed. Network data and network processing may be particularly difficult. It might be less costly and time consuming to adapt the network for the existing model calibration to the backcast year. Alternatively, if network impedance matrices are available for the backcast year, it may be possible to use the impedance data to validate travel model components such as trip distribution and mode choice.

Census Data

Census journey to work data from the 1990 and 2000 can provide valuable backcast validation data, especially for home-based work trip distribution or primary work location choice. Past Census data can also be valuable for temporal validation of population synthesis.

Historical Travel Model Applications

If travel survey data for a backcast year are unavailable, useful backcast validation data can be summarized from previous model applications, especially applications performed in support of previous model validation efforts. As noted for previous travel survey data, it may be difficult to process network data using existing modeling software. Thus, it might be necessary to restore previous releases of the modeling software.

Written Reports

Written travel model validation or model application reports for backcast years might provide summary tables and data necessary for a backcast model validation. It might be necessary to adjust existing year networks and socioeconomic data to reflect the backcast year.

Sources of Forecast Validation Data

As mentioned previously, a forecast validation describes an application of the travel model for any year after the model calibration year. A few locations may have results from a new travel survey available but alternative sources such as the census or the American Community Survey (ACS) are more likely sources for validation data.

Travel Survey Data

In some cases, new travel survey data may be available for model validation. While it is likely that the new model validation data would be used for an update of the travel model, that effort may be scheduled for a future time. As an example, a version of the Puget Sound Regional Council (PSRC) regional travel model was estimated in 2000 based on 1999 household survey data. The primary validation for that model was for the year 2000. PSRC collected a new household travel survey (more than 4,000 households) in 2006 to support their development of an activity-based model. However, PSRC also used the 2006 data to validate the model parameters estimated in the 2000.

American Community Survey

The initiation of the ACS provides the opportunity to perform forecast temporal validations. The decennial Census of Population and Housing collects data about the number of people residing in the United States and their relationship within a household, age, race, Hispanic origin (ethnicity), and sex. It also collects information about the number, occupancy status, and tenure (ownership status) of the nation's housing units. In the censuses of 1980, 1990, and 2000, information about topics such as income, education, employment status, disability status, housing value, housing costs, and number of bedrooms were asked on the "long form." Since there is no long form associated with the censuses starting in 2010, data on these topics will come from the ACS.

Instead of collecting data from about one in every six households once every 10 years, as with the decennial census long form, the ACS samples about one in every 40 addresses every year, or 250,000 addresses every month. This allows the Census Bureau to produce data every year rather than every decade. For areas with large populations (65,000 or more), survey estimates are based on 12 months of ACS data. For all areas with populations of 20,000 or more, the survey estimates are based on three years of ACS data. The Census Bureau is planning to produce estimates for all areas, down to the census tract and block group levels, based on five years of ACS data. The U.S. Census plans to release more ACS data each year as shown in Table 10.1.

Table 10.1 ACS Data Releases

Data Product	Population Threshold	Planned Year of Release				
		2009	2010	2011	2012	2013
1-year Estimates	65,000+	2008	2009	2010	2011	2012
3-year estimates	20,000+	2006-2008	2007-2009	2008-2010	2009-2011	2010-2012
5-year Estimates	All areas[a]	–	2005-2009	2006-2010	2007-2011	2008-2012

Source: U.S. Census Bureau.

[a] Five-year estimates will be available for areas as small as census tracts and block groups.

The following are some of the data that may be useful for travel model validation available in the ACS:

- Demographic Characteristics:
  - Age;
  - Sex; and
  - Relationship to householder.
- Economic Characteristics:
  - Income;
  - Labor force status;
  - Industry, occupation, and class of worker;

  - Place of work and journey to work;
  - Work status last year; and
  - Vehicles available.
- Financial Characteristics:
  - Tenure (owner/renter);
  - Housing value;
  - Rent; and
  - Selected monthly owner costs.
- Journey-to-Work Characteristics:
  - Location of most frequent workplace last week;
  - Mode of travel to work usually used last week;
  - Number of people in vehicle if auto, truck, or van used to get to work;
  - Normal departure time for work last week; and
  - Normal home-to-work journey time last week in minutes.

10.1.3 Temporal Validation Tests

Temporal validation tests will, of course, depend on the validation data available for the backcast or forecast year. While it may not be possible to test each of the model components for the backcast or forecast year, if the data are available, model component validation tests should be performed in addition to an overall system test comparing modeled to observed traffic counts and transit boardings. The model component tests can support the results of overall system tests. For example, if a reasonable match between modeled and observed trip generation rates for a backcast year can shown in addition to a reasonable match between modeled and observed traffic volumes, a "better" temporal validation can be claimed. Such results should provide more confidence in travel forecasts produced using the model.

Results from temporal validation tests of model components based on survey data from two points in time are not likely to be as good as those for a typical model validation. The differences may be caused by differences in survey or sampling methods, random error associated with surveys, and changes in travel behavior over time (which the model may or may not be sensitive to).

10.2 Sensitivity Testing

Temporal validations such as forecasting or backcasting are important for systemwide model validations. However, if there are limited population, employment, land use, or transportation system changes between the calibration and validation years, the temporal validation will provide little information regarding the sensitivity of the model.

Sensitivity testing can provide important information for model assessment. Sensitivity tests are not designed to tell whether the travel model is "correct" but, rather, to provide information about the overall behavior of the model. Sensitivity testing will help reduce unpleasant "surprises" that can occur when a forecast for a future year or an alternative does not produce expected results. A well structured sensitivity testing program provides travel modelers the opportunity to focus on the big picture of determining the overall reasonableness of the model in preparation for producing forecasts for specific studies.

Sensitivity tests for trip-based models should consider at least two major types of changes: land use changes and system changes. For activity-based models, analyzing the impact of a major policy change should also be performed. Simple examples would be changes in forecasts over time (e.g., 2005 to 2030), across alternatives (e.g., a corridor with and without a major new transit guideway project in 2030), and across policies (e.g., with and without peak-period congestion pricing charges in a Central Business District).

Since the sensitivity tests will be producing forecasts for alternatives that do not currently exist, it is impossible to tell what results are correct. Several options for determining the reasonableness of the resulting forecasts are outlined below:

- Establish expected outcomes via a panel of experts – This approach can be used for any sensitivity test. A panel of transportation planners and model users can be established. The existing conditions and changes being tested with the travel model should be described to the panel. The panel would then be asked to develop "forecasts" of the expected results based on their best collective judgment. The forecasts should be developed with as much specificity as possible. After forecasts are made with the model, the results should be summarized and compared to those from the expert panel. The results should be summarized and explained, especially for cases where the forecast results from the model differed from the results expected by the panel.

- Analyze the components of change for the results – This analysis approach is based on a detailed comparison of the forecast results for the alternative being tested to the base case.

- Compare to an existing forecast for the same alternative – Most regions developing new travel models are refreshing or replacing existing travel models. It is likely that a future year forecast exists for the current Regional Transportation Plan (RTP) for the region. Rerunning the existing RTP using the new travel model will demonstrate the sensitivity of the travel model to long-term population, employment, and transportation network changes. In addition, since the component results from the existing travel model for the RTP should be available, the differences in the sensitivity of the updated travel model could be compared and contrasted to the previous model.

11.0 Validation Documentation

11.0 Validation Documentation

Model validation results should be well documented in order to provide users of the travel forecasts the information they need to establish their confidence in the models. The model documentation should cover the limitations of the models as well as the capabilities of the models. If the model limitations and portions of the model that have not been validated are documented, users of the forecasts can reasonably assess the level of confidence they place in the forecasts. Understanding that a model cannot be used to test a specific issue or policy can, ultimately, lead to increased trust in the travel model when used for analyses for which it has been validated.

Model validation documentation has often been included as a chapter in model estimation and calibration documentation or interspersed in the various chapters of the model development documentation. This practice can minimize the importance of model validation, make the validation information difficult to find, or confuse the concepts of model estimation, calibration, and validation. For these reasons, the development of a standalone model validation document is recommended.

11.1 Executive Summary for Nonmodeler Users of Forecasts

An executive summary should contain sufficient information for the readers to become acquainted with the full report without reading it. The summary should contain:

- A statement of the purpose and need for the validation;
- An overview of the validation process, information on the validation data;
- A summary of the validation results;
- A summary of model strengths and weaknesses; and
- Information regarding the types of studies for which the model is valid and for which it should not be used.

The summary of validation results might be presented via a table of the validation tests performed for the various model components along with qualitative assessments of the results. The executive summary should avoid focusing mainly on traffic and transit assignment results and statistics.

Other executive summary information may include:

- Model area background information with maps of the region and major transportation facilities;
- Modeling "philosophy" overview describing the purpose and process of the travel model;
- Modeling process overview detailing the travel demand forecasting process and steps in terms that nonmodeler users can understand;
- Model development section summarizing the development, calibration, and validation of the model;
- Recent model enhancements and improvements; and
- Overview of the report and an explanation of how the validation summaries are reported.

The overall objective for the executive summary is to summarize the report findings so that both experienced model users and nonmodelers understand the usefulness and the limitations of the model.

■ 11.2 Component Validation

Chapters 3 through 9 of this manual describe the model components typically present in traditional trip-based and emerging activity and tour-based travel models. Model validation documentation should include sections for each of the model components even if validation tests were not performed for a component. Information that a model component has not been validated is crucial to assessing the overall model validity.

Validation documentation should also discuss the variables included in the model and how those variables influence the results. For example, mode choice documentation might note that auto operating costs are included in the model and that those costs represent items such as fuel costs, fuel efficiency, other out of pocket costs. This discussion might take place even if an explicit validation of model sensitivity to the variable has not been performed.

Throughout this manual, an effort has been made to clearly define model estimation, model calibration, and model validation. Continuing with this distinction suggests that model component validation should focus only on the validation tests performed and validation results obtained. Model adjustments and corrections necessary to obtain the validation results desired should be covered in model estimation and calibration documentation. Documenting model adjustments and corrections (a model calibration task) in the validation documentation introduces the risk that a future model user will not apply the proper model; documenting the information in both locations is acceptable, but introduces the difficulty of maintaining documentation consistency.

11.3 Model System Validation

The documentation of the travel demand forecast model components should be followed by a section summarizing the overall model system validation. The end results of the travel demand modeling process are generally considered to be highway assignments and transit ridership. In the past, "proof" that a model is valid has generally been provided by satisfying specific statistical standards such as obtaining an R^2 of 0.89. However, satisfying such a standard is neither necessary nor sufficient to prove that a model is valid.

The above should not be interpreted as dismissing the importance of reporting statistical matches of observed traffic volumes and transit ridership. These statistics may be quite useful in assessing how travel forecasts should be interpreted for project design. Care should be use to ensure that validation statistics are not overstated. For example, a systemwide R^2 of 0.95 suggests that 95 percent of the variation in traffic on facilities is "explained" by the travel model. However, a major contributor to such a statistic is the fact that higher-level facilities with more capacity receive more traffic than lower-level facilities. More informative statistics might be the R^2 values for facilities stratified by area type and facility type or by capacity or observed volume ranges.

11.4 Model Sensitivities

Sensitivity testing should be presented in a manner that allows the user to understand the impacts of changes in model inputs on forecast results. Sensitivity testing might present true validation results if it results from backcasting or forecasting using a model calibrated using data from a different year. Alternatively, the sensitivity testing might simply provide information on how the travel forecasts are impacted by changes in model inputs.

For New Starts forecasts, the FTA uses the concept that the travel forecasts should "tell a coherent story." For New Starts, this concept relates to the model structure and parameters and how well they describe how people behave in relation to their transportation options. This concept can be easily extended to the documentation of sensitivity testing results.

11.5 Documenting the Limits of the Model Validity

The model validation report should include a section reporting on the limits of the model validity. Travel demand models are often applied to provide forecasts for issues outside the scope or purview of the model. Documenting model limitations and providing guidance on model applications can be a valuable resource for end users.

The validation documentation should also include a section on the high and low variable values used in the model development. Such an approach would help future model users to avoid the use of model parameters that lie outside the range of the validated data used to develop the model.

11.6 Documenting Next Steps in Model Development/ Calibration/Validation

The model validation report is a primary document used to communicate information regarding the travel demand forecasting model. For all intents and purposes it is an infomercial for the model set; it details the model set, provides guidance on its use, and can be valuable for planning and prioritizing the next steps required for model development. This section can include:

- Future work plan elements or suggestions regarding which model components should be updated next;
- Future data collection efforts and surveys;
- Estimated schedule for model development and model validation; and
- Any areas of emerging research of the model development that might impact the users and stakeholders.

NOTICE

This report is being distributed through the Travel Model Improvement Program (TMIP).

FHWA-HEP-10-042

FHWA-HEP-10-042

Federal Highway Administration

TMIP

Travel Model Improvement Program

TMIP is funded by the Federal Highway Administration's Office of Planning, Environment and Realty's Surface Transportation Environment and Planning Cooperative Research Program (STEP).