

Technical Report Documentation Page

1. Report No. FHWA/TX-06/0-4713-1		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle SECURING TRUCK DATA FOR TEXAS				5. Report Date October 2004; Revised January 2006	
				6. Performing Organization Code	
7. Author(s) Jolanda Prozzi, Akshay Mani, Rob Harrison				8. Performing Organization Report No. 0-4713-1	
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3208 Red River, Suite 200 Austin, TX 78705-2650				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. 0-4713	
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, TX 78763-5080				13. Type of Report and Period Covered Technical Report: 9/1/2003-8/31/2004	
				14. Sponsoring Agency Code	
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration. Project Title: Development of Sources and Methods for Securing Truck Travel Data in Texas					
16. Abstract Freight transportation is critical to the economic prosperity of any region. Associated intercity and interstate truck flows, however, have an important impact on mobility, system performance, the safety of the road network, and the funding of the state-maintained road infrastructure. With the passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, increasing interest in freight modeling has emerged within statewide planning efforts, particularly the evaluation of the current and future freight transportation capacities necessary to ensure freight mobility. Although freight demand models are thus starting to emerge as tools to inform transportation policies, a critical challenge in the development of these models remains—namely, insufficient and inferior quality freight data. In August 2003, the Texas Department of Transportation (TxDOT) contracted with the Center for Transportation Research at The University of Texas at Austin to recommend a robust methodology to TxDOT planners for collecting and maintaining intercounty and interstate truck travel data in a format that can be used in the Statewide Analysis Model (SAM). This research has culminated in this document.					
17. Key Words truck data, statewide planning, freight surveys, freight modeling, truck forecasting				18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov .	
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified		21. No. of pages 136		22. Price



Securing Truck Data for Texas

Jolanda Prozzi
Akshay Mani
Rob Harrison

CTR Research Report:	0-4713-1
Report Date:	October 2004; Revised January 2006
Research Project:	0-4713
Research Project Title:	Development of Sources and Methods for Securing Truck Travel Data in Texas
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
3208 Red River
Austin, TX 78705

www.utexas.edu/research/ctr

Copyright © 2006
Center for Transportation Research
The University of Texas at Austin

All rights reserved
Printed in the United States of America

Disclaimers

Authors' Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Notice: The United States Government and the State of Texas do not endorse products or manufacturers. If trade or manufacturers' names appear herein, it is solely because they are considered essential to the object of this report.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Research Supervisor: Rob Harrison

Acknowledgments

The authors wish to acknowledge the involvement and direction of the Texas Department of Transportation (TxDOT) project monitoring committee—specifically, Project Director Gurjeet Gill (TPP), Project Coordinator Augustin De La Rosa (IRO), and Project Monitoring Committee Members Charles Davidson (Austin), Charlie Hall (TPP), Janie Bynum (TPP), Jim Neidigh (TPP), Eric Oeding (TPP), Charles Airiohuodion (Houston), and Mike Sims (NCTCOG). Special thanks is also owed to Akshay Mani (CTR) for his review of available freight disaggregation methodologies and his time spent in developing the multinomial logit modeling approach for estimating county-level truck data for Texas, Ka Yan Lo (CTR) for data analysis, and Carleton Wong (CTR) for his work on the survey approaches. Finally, thanks are owed to Yulawnda Hines (CTR) and Maria Saenz-Farias (CTR), who provided administrative and graphical assistance in the production of this document.

Table of Contents

1. Introduction.....	1
1.1 Background	1
2. Available Freight Data Sources	5
2.1 Data Used by State Departments of Transportation.....	5
2.2 Available Freight Databases	8
2.3 Concluding Remarks.....	13
3. Populating Statewide Models Using Publicly Available Sources	15
3.1 Statewide Analysis Model’s Data Requirements and Structure	15
3.2 Review of Freight Flow Disaggregation Methods.....	16
3.3 Concluding Remarks.....	20
4. Proposed Methodology for Populating Statewide Analysis Model Database Using Commodity Flow Survey Data.....	21
4.1 Multinomial Logit Model.....	21
4.2 Calibration of Multinomial Logit Model to Estimate Truck Flows for Texas.....	24
4.3 Truck Flow Disaggregation to the County Level.....	31
4.4 Results Of Model Calibration	38
4.5 Calculation Of County-Level Truck Flows	42
4.6 Strengths and Limitations of the Modeling Methodology	48
4.7 Comparison of Model Output with Reebie TRANSEARCH Database.....	50
4.8 Concluding Remarks.....	52
5. Truck Travel Survey Methods.....	53
5.1 Truck Travel Survey Methods	53
5.2 Recommended Truck Data Collection Approaches.....	56
5.3 Truck Carrier Participation	68
5.4 Cost Effectiveness.....	72
5.5 Concluding Remarks.....	73
6. Future Opportunities for Collecting State Truck Data.....	75
6.1 Security Initiatives	75
6.2 Proposed National Freight Data Program	76
6.3 Real Time Truck Data Collection Methods	76

6.4 Concluding Remarks.....	80
7. Freight Forecasting.....	81
7.1 Statewide Freight Demand Forecasting Techniques.....	81
7.2 State-Of-The-Practice in Statewide Truck Forecasting	84
7.3 Proposed Truck Traffic Forecasting Procedure for The Statewide Analysis Model	88
7.4 Concluding Remarks.....	95
8. Conclusions and Recommendations.....	97
8.1 Freight Data Used by State Departments of Transportation	97
8.2 Available Freight Data Sources	97
8.3 Populating State Transportation Models Using Public Data.....	98
8.4 Proposed Approach for Estimating State Truck Data	98
8.5 Truck Travel Survey Methods	100
8.6 Future Opportunities for Collecting State Truck Data.....	101
8.7 Freight Forecasting Techniques	102
8.8 Concluding Remarks and Recommendations	103
9. References.....	105
Appendix A: Available Freight Data Sources	109
Appendix B: Mississippi Intermodal Freight Model	119
Appendix C: List of Acronyms	121

List of Figures

Figure 1.1 Four Stages of Commodity-based Truck Demand Modeling.....	2
Figure 3.1 Factors Affecting Freight Transportation Demand	17
Figure 4.1 Truck flows of food commodities from Harris County to Colorado.....	43
Figure 4.2 Truck flows of food commodities from Colorado to Harris County.....	44
Figure 4.3 Truck flows of food commodities from Harris County to Dallas County.....	46
Figure 5.1 Proposed Truck Travel Questionnaire.....	63

List of Tables

Table 2.1 Freight Planning Programs and Freight Data Sources Used by U.S. States	6
Table 2.2 Shipment-/Commodity-Based Data Sources	10
Table 2.3 Transportation-/Mode-Based Data Sources.....	10
Table 2.4 Commodity/Industry Classifications	11
Table 2.5 Origin–Destination of Truck Shipments.....	12
Table 2.6 Truck Shipment Characteristics.....	12
Table 2.7 Modal Characteristics	13
Table 3.1 Aggregated Commodity Categories Included in SAM.....	16
Table 4.1 Outputs of the OLS Regression Analysis for the Production Flow Distribution Model	40
Table 4.2 Outputs of the OLS Regression Analysis for the Attraction Flow Distribution Model	42
Table 4.3 Paired Sample t Test Results for the Texas County-to-State Truck Flow Comparison	51
Table 4.4 Paired Sample t Test Results for the State-to-Texas County Truck Flow Comparison	51
Table 4.5 Paired Sample t Test Results for the Texas County-to-County Truck Flow Comparison	52
Table 5.1 Summary of Truck Travel Survey Methods	54
Table 5.2 Explanation of the Data Fields.....	64
Table 5.3 Seasonal Weight Factor	67
Table 5.4 Site Specific Weight Factor	68
Table 5.5 Potential Sources to Identify Trucking Companies Operating on Texas Roads.....	69
Table 6.1 Attributes Captured by Current ITS Technologies	79
Table 7.1 Explanatory Variables to Model Commodity Productions and Attractions	92
Table 7.2 Average Truck Payload Factors by Commodity Group	95

1. Introduction

Freight transportation is critical to the economic prosperity of any region and quality of life. Associated intercity and interstate truck flows, however, have an important impact on traffic volumes, the mix of traffic, and experienced levels of congestion on the state-maintained infrastructure. The federal government recognized the importance of an efficient freight transportation system to the economic vitality of a state by passing the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991, which emphasized the inclusion of freight planning as a critical component of the statewide transportation planning process. The objectives of the ISTEA include enhancing accessibility to ports, airports, intermodal transportation facilities, and major freight distribution centers, and the efficient movement of commercial vehicles for domestic and international trade. The ISTEA was reauthorized as the Transportation Equity Act for the twenty-first century (TEA-21) in 1998, which reinforced the mandate for more comprehensive freight planning programs by state and metropolitan planning agencies. The passage of the ISTEA and TEA-21 have thus resulted in an increasing interest in freight modeling within statewide planning efforts, and particularly the evaluation of the current and future freight transportation capacities necessary to ensure freight mobility.

1.1 Background

A number of factors influence the demand for truck transportation, including the economy, land use patterns (e.g., the location of industrial centers), supply and logistics strategies, trade and transportation agreements, and legislation (environmental, tax, and transportation policies). These various factors interact to translate finally in truck volumes on the state's infrastructure. An understanding of these factors, how they change, and how they interact is necessary to model current and estimate future impacts on truck traffic volumes.

A review of the literature on freight transportation demand models revealed that freight demand studies tend to fall into one of two categories: commodity-based analyses, focusing on the flow of goods, or truck traffic (vehicle-based) analyses, focusing on the flow of vehicles. The former have been applied at the regional and state level, whereas truck models have been applied on an urban scale (Donnelly, undated).

Commodity-based models emphasize the producers and consumers of goods. These models estimate the demand for interzonal commodity movements as a function of the socioeconomic characteristics of zones. Future freight flows are estimated by forecasting the commodity flows as a function of the future socioeconomic characteristics of zones derived from standard federal data sources developed by the Bureau of the Census and the Bureau of Economic Analysis (BEA). The typical steps involved in commodity-based truck demand modeling are as follows (see Figure 1.1).

- Commodity flow generation. The first step of the modeling process is to develop mathematical models representing the commodity productions and attractions in each zone as a function of the zonal socioeconomic characteristics. Future commodity productions and attractions in each zone are then computed from the calibrated models using forecasted socioeconomic data.

- **Commodity flow distribution.** Using the current interzonal commodity flow matrix, a commodity flow distribution model is calibrated to best represent the observed values. The forecasted commodity productions and attractions in each zone are then distributed to generate forecasted interzonal commodity flows.
- **Conversion to truck trips.** The future demand for freight transportation is expressed in terms of the number of truck movements on the system. Forecasted interzonal commodity flows (shipment tonnage) are therefore converted to truck trips using average truck payload factors for each commodity category.
- **Truck traffic assignment.** The final step of the modeling process is to assign the interzonal truck flows to the highway network using, for example, standard shortest path traffic assignment techniques.

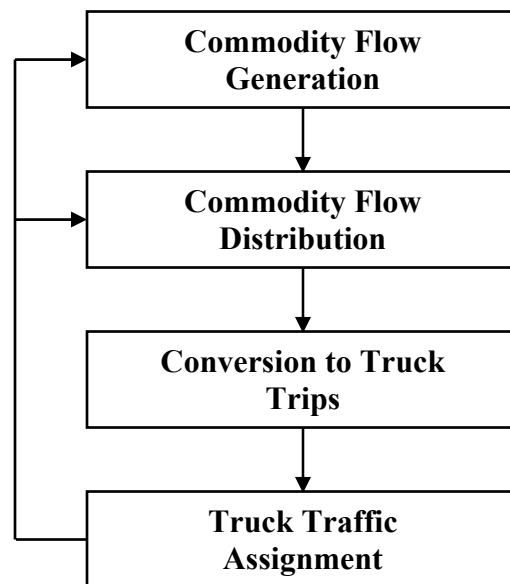


Figure 1.1 Four Stages of Commodity-based Truck Demand Modeling

Vehicle-based models have sequential steps for trip generation, distribution, mode choice and network assignment—analogue to the person travel models (Donnelly, undated). Vehicle-based models do not explicitly model the demand for commodity flows and therefore directly model the movement of trucks on the highway system. In the vehicle-based approach, truck flows are estimated either by using mathematical models as functions of zonal (i.e., land use) variables or by empirical methods in which truck flows are estimated from forecasted passenger flows. Typical steps involved in vehicle-based modeling are as follows.

- **Truck flow generation.** In this step, mathematical models are calibrated to represent the total truck trips produced and attracted in each zone as a function of socioeconomic and land use variables of zones. Future truck flows are estimated

from the calibrated models using forecasted values of the socioeconomic and land use variables.

- Truck trip distribution. Using the derived interzonal truck trip matrix, a truck trip distribution model is calibrated to best represent the observed values. Most planners rely on traditional travel demand modeling software packages, such as EMME-2 and Tranplan, to distribute truck trips.
- Truck traffic assignment. Interzonal truck flows are then assigned to the highway network using, for example, shortest path traffic assignment techniques.

Although freight demand models are thus emerging as tools to inform transportation policies, a critical challenge in the development of these models remains—namely, insufficient and inferior quality data. Most states that have conducted statewide freight modeling seem to have relied on the commercial Reebie TRANSEARCH database, in part because this is currently the only database that captures most of the variables needed for freight modeling. In August 2003, the Texas Department of Transportation (TxDOT) contracted with the Center for Transportation Research at The University of Texas at Austin to recommend a robust methodology to TxDOT planners for collecting and maintaining intercounty and interstate truck travel data in a format that can be used in TxDOT's Statewide Analysis Model (SAM). The research effort is documented in this research report, which is structured as follows. Chapter 2 reviews the freight data used by state departments of transportation in their modeling efforts and the freight databases available to transportation planners. Chapter 3 briefly highlights the data requirements of the freight component of the SAM and discusses a disaggregation approach that can be used to estimate county-to-county flows using publicly available commodity flow and economic data. Chapter 4 proposes a methodology to construct a truck database for the SAM from the publicly available Commodity Flow Survey data. The modeling steps involved in estimating disaggregated truck data for Texas using the multinomial logit model approach are discussed in detail. Chapter 5 provides an overview of the available primary freight data collection methods that have been used and discusses two data collection approaches that show the most promise in providing TxDOT with the needed information for populating the truck database required for the SAM. Chapter 6 highlights a number of national initiatives for collecting data that might be accessible to state departments of transportation in the future. Chapter 7 provides an overview of the various freight-forecasting techniques available, ranging from simple growth factors for short-term forecasts to more complex models for long-term forecasts. In addition, the state-of-the-practice in statewide truck forecasting is discussed before the chapter concludes with two recommended freight-forecasting procedures for the SAM. Finally, Chapter 8 presents the main conclusions and recommendations of this research.

2. Available Freight Data Sources

Although the economic benefits of freight transportation are seldom disputed, the challenge lies in disaggregating freight transportation demand to truck flows that can be assigned onto a state's transportation network to facilitate planning. State departments of transportation require disaggregated freight flows to accomplish the following goals:

- provide a clear picture of freight movements on a state's transportation system;
- determine the impact of freight on a state's road infrastructure (e.g., bridges and pavements) and the implications in terms of funding;
- evaluate strategies for improving freight mobility;
- forecast system performance;
- mitigate impacts of truck traffic on general mobility;
- determine the impacts on air quality;
- ensure effective land use planning;
- evaluate economic development impacts; and
- improve the safety and security performance of the road network.

The evaluation of current and future freight transportation capacity is thus critically contingent on the availability of accurate data and sound models to ensure informed decisions. The objective of this section of the report is to provide an overview of the freight data used by state departments of transportation in their modeling efforts and the freight databases available to transportation planners.

2.1 Data Used by State Departments of Transportation

State departments of transportation rely (a) predominantly on traffic and classification counts conducted, (b) on the limited data compiled and published by federal agencies for aggregate analysis, (c) on one of the private commercial sources of data related to freight movements, or (d) on collected original data. The principal commercial data source is the Reebie TRANSEARCH database (see Table 2.1 for a summary of the freight planning models and studies that various state departments of transportation have embarked upon and the sources of freight data that were used in these efforts). The Reebie TRANSEARCH database is a unique source of detailed freight data that is available for purchase. The data sources used to compile the database are proprietary, and the assumptions used to estimate and forecast the data are not disclosed. It is thus not possible to easily verify the accuracy and reliability of the data. This lack of transparency regarding the sources of the data, the methodology used, and the assumptions made have raised several questions about the validity and reliability of the Reebie TRANSEARCH database. The database, however, remains the most often used database for statewide analysis of freight movements.

Table 2.1 Freight Planning Programs and Freight Data Sources Used by U.S. States

State	Freight Planning Models/Studies	Freight Data Sources			
		TRANSEARCH	Commodity Flow Survey	Surveys/	Traffic/Class Counts
California	Intermodal Transportation Management System	√	-	√	√
Colorado	Eastern Colorado Mobility Study	√	-	-	√
Florida	Florida Intermodal Statewide Highway Freight Model	√	-	√	√
Indiana	Indiana Commodity Flow Model	-	√	-	√
Iowa	Iowa Statewide Transportation Planning Model	√	-	√	√
Kansas	Kansas Statewide Agricultural Model	-	-	√	√
Maine	Maine Integrated Freight Plan	√	-	-	√
Massachusetts	Massachusetts Truck Model	-	√	-	√
Michigan	Michigan Statewide Truck Model	-	√	√	√
Minnesota	Minnesota Regional Freight Flow Study	√	√	√	√
Mississippi	Intermodal Freight Transportation Planning Methodology	-	√	-	√
Ohio	Freight Impacts on Ohio's Roadways	√	-	-	√
Oklahoma	Oklahoma Freight Movement Model Development	-	√	-	√
Oregon	Oregon Integrated Statewide Model	-	-	-	√
Texas	Statewide Analysis Model (SAM)	√	-	√	√
Vermont	Vermont Statewide Freight Study	√	-	√	√
Virginia	Virginia Statewide Intermodal Freight Planning Methodology	√	-	-	√
Washington	Washington State Freight Truck Origin and Destination Study	-	-	√	√
Wisconsin	Wisconsin Statewide Model	√	-	-	√

The availability of statewide truck flow data remains as a major challenge for states in developing freight demand models. In addition, interviews with these states and a detailed review of the literature review suggests that few states rely on the collection of primary freight data, partially because this can be a costly and time-consuming process at the state level. The exception has been Washington State's Eastern Washington Intermodal Transportation Study (EWITS). The objective of the EWITS was to conduct a statewide origin–destination truck survey through direct intercept surveys of truck drivers in 1993/94 (see the text box). The criteria employed in deciding the survey methodology were as follows.

- “Data collected should provide statistically reliable information on truck characteristics and commodity flows for all major Washington highways.
- The sample size should be large enough to provide useful freight and goods movement information for major transportation planning subregions as well as the state as a whole.
- Information, where available, should be developed over a continuous 24-hour period in each of the four seasons of the year.” (Jessup, Casavant, and Lawson, 2004)

On the basis of these criteria, roadside intercept interviews were judged as the most effective method for collecting truck travel data for Washington State.

Washington State's Eastern Washington Intermodal Transportation Study (EWITS)

A total of 28,000 truck drivers were interviewed by over 300 interviewers at twenty-seven different locations (Jessup, Casavant, and Lawson, 2004), including twenty-one Washington State Patrol Weigh Stations, three border locations with Canada, and the Oregon Port of Entry at Umatilla (Gillis, Jessup, and Casavant, 1995). To the extent feasible, interviews were conducted for 24 hours at each site on Wednesdays in each of the four seasons. The survey instrument sought information on “*time-of-day movements, vehicle configuration, trucking company location, origin and intended destination, cargo type, commodity, hazardous material, loaded weight, empty weight, owner of truck, type of destination facility and type of origin facility, specific route, and several other characteristics*” (Jessup, Casavant, and Lawson, 2004). Many of the questions could be answered through visual observation, which allowed for the completion of the questionnaires within 3 minutes. Driver participation was excellent, with more than 96 percent of the truck drivers agreeing to an interview when asked to participate (Gillis, Jessup, and Casavant, 1995).

Finally, a few states have developed procedures to disaggregate truck flows from publicly available data sources, such as the Commodity Flow Survey. The county-level freight disaggregation procedure is generally based on county socioeconomic data gathered from standard data sources such as county business patterns (CBPs) and county population estimates published by the U.S. Bureau of the Census as well as the IMPLAN economic impact modeling databases developed by the Minnesota IMPLAN Group (MIG), Inc. Chapter 3 of this report reviews the freight flow disaggregation methods that have been documented in the literature.

2.2 Available Freight Databases

A number of private and public freight databases exist (see Appendix A for a summary of the available sources and the most important variables recorded in these databases). These freight data sources, however, differ in terms of the following:

- scope and structure;
- coverage of commodity characteristics;
- coverage of origin/destination characteristics;
- coverage of shipment characteristics; and
- coverage of transportation characteristics.

Of the more than fifty private and public freight data sources reviewed as part of the literature review, only six capture some of the variables of interest to TxDOT —such as truck volume (tonnage, value, and number of loads), origins and destinations, commodities, truckload or less-than-truckload —at various levels of detail (see the text box). These data sources are as follows.

- Commodity Flow Survey (CFS)
- Reebie Associates' TRANSEARCH
- Freight Transportation and Logistics Service
- Transborder Surface Freight Database
- North American Trucking Survey (NATS)
- Vehicle Inventory and Use Survey (VIUS)

For additional information about the objectives, survey methodologies, variables captured, assumptions, and limitations of each publicly available and commercial freight data source, the reader is referred to a document entitled *State-Of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S.*, which was prepared as part of this TxDOT research project.

Available Freight Data Sources

Commodity Flow Survey. Publicly available database developed by the U.S. Bureau of the Census and the Bureau of Transportation Statistics every five years as part of the Economic Census. Captures shipment data from manufacturing, mining, wholesale, and selected retail and service establishments. The shipment data include distance distributed and origin–destination flows (interstate and intrastate by National Transportation Analysis Regions) by commodity type, mode, shipment size, and value.

Reebie TRANSEARCH database. Multimodal freight flow database that displays commodity tonnage and loads by mode and between origins and destinations at the county, business economic area (BEA), metropolitan area, and state or provincial level. The modes included are for-hire truckload, for-hire less-than-truckload, private truck, rail carload, rail/truck intermodal, air, and water. The database is fused from various commercial, public, and proprietary freight data sources, including data from trucking companies shared as part of a Motor Carrier Data Exchange program.

Freight Transportation and Logistics Service. A proprietary freight data source developed by DRI/McGraw-Hill Inc. that provides historical and forecasted data for commodity and modal traffic (barge, rail, and truck), cost, rate, and equipment demand. Detailed shipment weight data are provided for private and for-hire truckload and LTL shipments at the two-digit STCC level.

TransBorder Surface Freight Data. Publicly available freight data source developed by the Bureau of Transportation Statistics (BTS) under a contract with the U.S. Bureau of the Census. Captures North American trade by commodity type, mode, and geographic detail for U.S. exports to and imports from Canada and Mexico. Data can be downloaded from the Internet.

Vehicle Inventory and Use Survey (VIUS). Publicly available data source developed by the U.S. Bureau of the Census. Vehicle-based survey of licensed (registered) private and commercial trucks as of July 1 of each survey year. Captures information on vehicle ownership, body type (e.g., single unit, tractor-trailer), equipment type (e.g., engine type), leasing activity, truck configurations (e.g., number of axles), dimensions (e.g., length), capacity (e.g., gross vehicle weight), mileage, commodities transported, and operating characteristics (e.g., percentage of annual mileage by commodity type and range of operation).

North American Trucking Survey (NATS). Proprietary database containing information on a sample of predominantly long-haul truckload movements (e.g., city and state of origin and destination, shipment weight, and commodity as a three-digit STCC), operator characteristics (e.g., for-hire, private or owner-operator), and annual vehicle miles traveled (VMT). The database was first developed in 1993 for the American Association of Railroads (AAR) from roadside interview surveys of long-haul truck drivers. Data are often made available to federal and state agencies when requested.

Source: Mani and Prozzi, 2004

2.2.1 Scope and Structure

Available freight data sources vary considerably in terms of scope and structure. In general, though, these data sources can be categorized as either shipment-based or transportation-based databases. Shipment-based databases (see Table 2.2) contain records for individual shipments. The best-known shipment-based database is the Commodity Flow Survey database compiled by the Census Bureau in partnership with the Bureau of Transportation Statistics (NCHRP Report 388, 1997).

Table 2.2 Shipment-/Commodity-Based Data Sources

Public	Commercial/Proprietary
Commodity Flow Survey (CFS)	TRANSEARCH
Transborder Surface Freight Database	North American Trucking Survey

Transportation databases (see Table 2.3), on the other hand, cover primarily that portion of each trip movement made on a specific mode. Transportation-based databases can include modal origin–destination flows, point activity at transportation nodes, carrier profiles such as the Truck Inventory and Use Survey (TIUS) and the National Truck Activity and Commodity Survey (NTACS), and modal profiles. Modal profiles provide a more generalized overview of modal activity at a regional or national level, usually excluding carrier or network information. Databases such as Piers are considered transportation databases, but Piers contains some information on the actual origins and destinations of shipments (NCHRP Report 388, 1997).

Table 2.3 Transportation-/Mode-Based Data Sources

Public	Commercial/Proprietary
Vehicle Inventory and Use Survey	Freight Transportation and Logistics Service
State Estimates of Truck Traffic	

A number of databases, however, do not fall into either of these categories, such as those that capture point activity at origin or destination or those that relate to a specific commodity or market profile (NCHRP Report 388, 1997). Databases showing point activity at an origin and/or destination usually do not include any modal detail. Finally, the data contained in the TransBorder Surface Freight Database (collected by U.S. Customs for the purpose of collecting tariffs on imports) provide an invaluable source of information to monitor freight flows from Mexico and Canada into the U.S.

Commodity Coverage

The extent and level of commodity detail captured varies from one data source to the other, and is usually a reflection of the intended objectives of the database. For example, the trade flow databases typically use product-based classification systems such as the Harmonized Schedule of Foreign Trade (HS) and the Standard International Trade Classification (SITC), whereas transportation-oriented sources typically use classifications such as the Standard Transportation Commodity Codes (STCC) (see Table 2.4). On the other hand, certain node specific sources may classify freight solely on the basis of handling characteristics, such as bulk, container, or break-bulk (NCHRP Report 388, 1997).

Table 2.4 Commodity/Industry Classifications

Data Source	HS	SCTG	SITC	STCC	SIC	NAICS
CFS	-	√	-	-	-	-
TRANSBORDER	√	-	-	-	-	-
VIUS	26 standard commodity categories and 17 HAZMAT categories				-	-
TRANSEARCH	-	-	-	√	-	-
FTLS	-	-	-	√	-	-
NATS	-	-	-	√	-	-

2.2.2 Coverage of Origin/Destination Characteristics

In most databases, origins and destinations (O/Ds) are specified at a very aggregate level of detail, such as the state level (see Table 2.5). In the case of the transportation databases, the level of O/D detail varies considerably, from aggregate O/Ds at the state or international level to actual names and locations of shippers and consignees. In most cases, however, if O/D information is captured, the information is aggregated into Business Economic Areas (BEAs) or National Transportation Analysis Regions (NTARs), although the Piers database contains actual shipper and consignee names and locations (NCHRP Report 388, 1997).

Table 2.5 Origin–Destination of Truck Shipments

Data Source	Country		State	Province	NTAR	County	City/ Zip Code
	Import	Export					
CFS	-	√	√	-	√	-	-
TRANSBORDER	√	√	√	√	-	-	-
VIUS	-	-	-	-	-	-	-
TRANSEARCH	√	√	√	√ (Canada)	√	√	√ (Zip code)
FTLS	-	-	-	-	-	-	-
NATS	√	√	√	-	-	-	√ (City)

2.2.3 Coverage of Shipment Characteristics

Shipment characteristics are described by variables that record shipment volumes, weights, seasonality, and other factors (see Table 2.6). Total shipment value is usually available from the trade-related sources and is primarily measured at U.S. points of import or export. The Interstate Commerce Commission (ICC) Waybill sample records the number of carloads for each shipment (NCHRP Report 388, 1997).

Table 2.6 Truck Shipment Characteristics

Data Source	Weight	Volume	Value	Number of Shipments	Containerized Designation
CFS	√	-	√	√ (Containerized)	√
TRANSBORDER	√ (Imports)	-	√	-	√ (Imports)
VIUS	-	-	-	-	-
TRANSEARCH	√	-	√	√ (Loads)	-
FTLS	√	-	-	-	-
NATS	√	-	-	-	-

2.2.4 Coverage of Transportation Characteristics

Important transportation characteristics include modal coverage (see Table 2.7), equipment detail, routing detail, and carrier/service detail. Routing information is usually limited to origin and/or destination information for a specific mode, aggregated into regions. Some sources,

though, provide distance information (e.g., CFS), usually estimated from inferred routings (NCHRP Report 388, 1997).

Table 2.7 Modal Characteristics

Data Source	Truck Inventory	Employment Statistics	Operating Revenue/ Expenses	Operator Type	Physical Characteristics	Operational Characteristics
CFS	-	-	-	√	-	-
TRANS-BORDER	-	-	-	-	-	-
VIUS	√	-	-	√	√	√
TRANSEARCH	-	-	-	√	-	-
FTLS	√	-	√	-	-	-
NATS	-	-	-	√	-	√

2.3 Concluding Remarks

From the above discussion, it is obvious that a number of freight data sources exist that capture some of the variables of interest to TxDOT, although at various levels of detail. No single data source, with the exception of the Reebe TRANSEARCH database, contains all the data necessary to populate an intercounty and interstate truck travel database for Texas. In addition, a number of these data sources are incompatible, largely because of the following factors.

- Insufficient detail (in terms of sample sizes, assumptions, reliability, etc.) reported that prevents the comparison and combination of different data sources.
- Different protocols in assigning O/Ds of truck traffic. For example, some of the databases assign O/Ds on the basis of billing/documentation locations, whereas others assign the location where the information is recorded rather than the point of production or consumption. There is also the issue of whether the origin is the origin of the shipment or the listed address of the shipper and confusion about whether the origin or destination is the administrative port of entry/exit, where the paperwork is filed or the physical port of entry/exit, where the shipment physically enters or exits the US.
- Different commodity classifications used.
- Different assumptions to estimate data or deal with missing data.
- Different expansion factors and control totals used.
- Different procedures in assigning O/Ds for multiple load shipments (e.g., origins of production and destinations of consumption vs. truck trip origins and destinations).

- Different O/D definitions that do not directly correlate, such as Business Economic Area (BEA) definitions as opposed to state-based statistics.
- Different procedures used for data aggregation.

Given these incompatibilities, concerns exist about the eventual quality and reliability of a database that results from the “fusion” of different databases.

3. Populating Statewide Models Using Publicly Available Sources

Truck traffic is a vital component of the transportation system in Texas, accounting for a major percentage of the goods movements within, to, from and through the state. A number of factors, including the deregulation of the transportation sector in the 1980s, a rapid economic growth rate in the 1990s, the increase in U.S. international trade with Canada and Mexico associated with the North American Free Trade Agreement (NAFTA), and the adoption of just-in-time logistics and manufacturing practices have resulted in a significant increase in commercial truck traffic in Texas. According to the Freight Analysis Framework (FAF), trucks are expected to carry a staggering 62.6 percent of the tonnage and 75 percent of the value of freight shipments in Texas in 2020 (Federal Highway Administration, 1999).

The Texas Department of Transportation (TxDOT) has funded the development of a Statewide Analysis Model (SAM) to assess the flows of passengers and freight on the state-maintained roadways. The objective of the SAM is to provide a regional model for the state of Texas that focuses on intercounty travel patterns. This section of the report briefly highlights the data requirements of the freight component of the SAM and discusses a disaggregation approach that can be used to estimate county-to-county flows using publicly available commodity flow and economic data.

3.1 Statewide Analysis Model's Data Requirements and Structure

The freight component of the model uses county-to-county commodity data (tonnage and number of loads) captured in the Reebie TRANSEARCH database. Approximately 4,600 internal Traffic Analysis Zones (TAZs) are included in the SAM, as well as 142 external TAZs. The county-to-county truck tonnage is disaggregated to the TAZs using employment data. SAM can thus display the statewide truck traffic flows for eleven commodity categories (see Table 3.1) for a base and forecasted year (see Chapter 7 for a discussion on how truck traffic is forecasted in the SAM). An embedded TransCAD function assigns the truck tonnage data to the network.

The SAM thus requires disaggregated commodity truck tonnage and flow (e.g., number of loads) data for internal–internal, internal–external and external–internal truck flows in Texas. In particular, commodity tonnage and number of loads with the following origin–destination characteristics are needed:

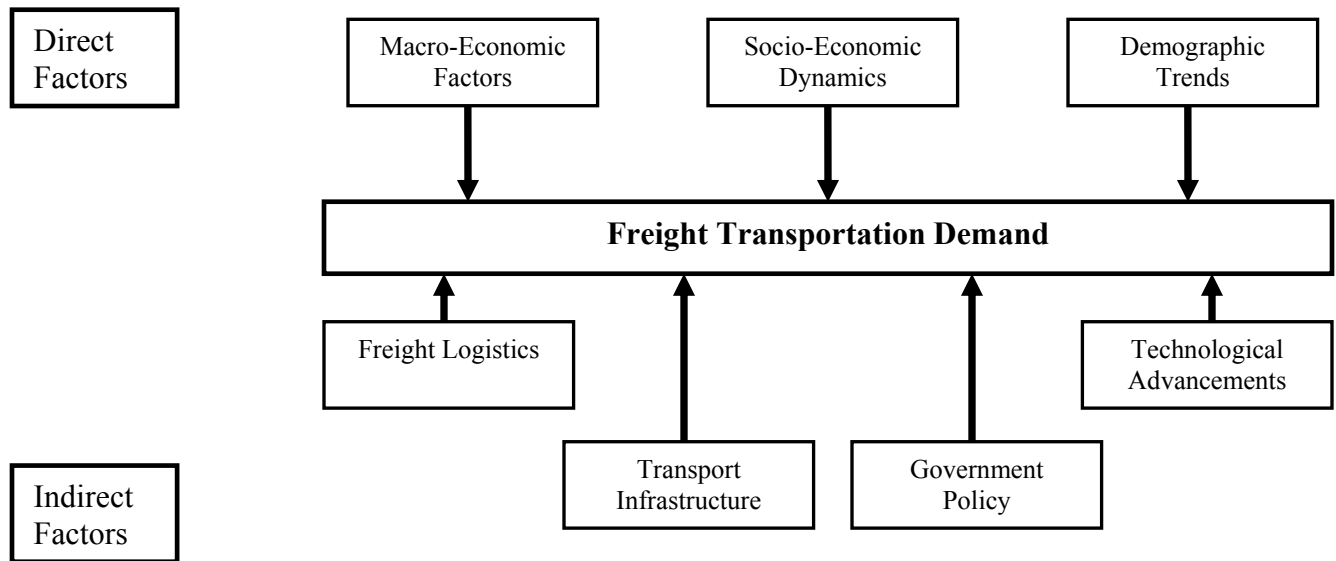
- Texas counties-to-states (internal–external),
- Texas county exports-to-Mexico/Canada (internal–external),
- States-to-Texas counties (external–internal),
- Texas county imports-from-Mexico/Canada (external–internal),
- Texas county-to-county flows (internal–internal), and
- Texas through flows (external–external).

Table 3.1 Aggregated Commodity Categories Included in SAM

Commodity Group	Commodity Categories
Agriculture	Live animals and live fish; cereal grains; other agricultural products; animal feed and products of animal origin, n.e.c.
Food	Meat, fish, seafood, and their preparations; milled grain products and preparations, and bakery products; other prepared foodstuffs and fats and oils; alcoholic beverages; tobacco products
Building Materials	Monumental or building stone; nonmetallic mineral products; base metal in primary or semifinished forms and in finished basic shapes; articles of base metal
Raw Material	Natural sands; gravel and crushed stone; nonmetallic minerals n.e.c.; metallic ores and concentrates; coal
Chemicals/Petroleum	Gasoline and aviation turbine fuel; fuel oils; coal and petroleum products, n.e.c.; basic chemicals; pharmaceutical products; fertilizers; chemical products and preparations, n.e.c.
Wood	Logs and other wood in the rough; wood products; pulp, newsprint, paper, and paperboard; paper or paperboard articles; printed products; furniture, mattresses and mattress supports, lamps, lighting fittings
Textiles	Plastics and rubber; textiles, leather, and articles of textiles or leather
Machinery	Machinery; electronic and other electrical equipment, components and office equipment; motorized and other vehicles (including parts); transportation equipment, n.e.c.; precision instruments and apparatus; miscellaneous manufactured products
Miscellaneous	Waste and scrap; mixed freight
Secondary	Warehouse and distribution; truck intermodal drayage; truck air drayage
Hazardous	Waste hazardous materials; hazardous materials and substances

3.2 Review of Freight Flow Disaggregation Methods

Freight demand is primarily driven by economic, socioeconomic, demographic, cost, and level of service factors (see Figure 3.1). Therefore, freight movements are primarily driven by commodity supply and demand and the indirect factors. Freight demand has been modeled as a function of the relative commodity production and attraction levels of regions and an appropriate impedance measure, which may represent the cost of transporting goods or the distance between the regions as a proxy for the transportation cost.



Source: Pendyala, Shankar, and McCullough, 2000

Figure 3.1 Factors Affecting Freight Transportation Demand

3.2.2 Available Economic and Freight Data

Economic data

County commodity productions can be estimated from county industry employment data associated with the respective commodities. County employment data are available from the County Business Pattern database developed annually by the Census Bureau. The County Business Pattern database captures detailed information about the total number of establishments, number of establishments by various employment size classes, and total employment in terms of the North American Industry Classification System (NAICS).

County commodity attractions can similarly be estimated as a function of the county population and intermediate industry employment. County population is a proxy for the personal consumption of commodities in the county, whereas intermediate industry employment is a proxy for the industrial consumption of commodities produced in the county. County population statistics are available from the County Population Estimates data published annually by the Census Bureau. Intermediate industry employment data can be obtained from the County Business Pattern database after identifying the commodities consumed by the intermediate industries from the U.S. input/output (I/O) accounts. The U.S. I/O accounts consist of detailed benchmark I/O accounts that are developed once every five years, although less detailed updates are available annually. The detailed I/O accounts are based on the U.S. economic census data collected by the U.S. Bureau of the Census. These I/O accounts provide detailed data on the level of production of goods and services, the level of commodity consumption by each industry

sector, the contribution of each industry sector to the gross domestic product, and the value added as presented by the contribution of primary inputs to the final industry output.

Finally, an important tool used by public and private agencies for in-depth economic analysis of states, counties, and multicounty regions in the U.S. is the IMPLAN (IMPact analysis for PLANning) economic impact modeling system developed by the Minnesota IMPLAN Group, Inc (MIG, Inc). The IMPLAN databases provide comprehensive information on the institutional demand of commodities, total industry output, employment and value added for each county in the U.S., along with state and national totals.

Freight Data

Among the publicly available freight data published on a continual basis, the Commodity Flow Survey (CFS) is potentially the most useful for generating disaggregate truck flows in Texas pending a decision from TxDOT not to purchase the Reebie TRANSEARCH database. The CFS captures shipment data from a sample of businesses in the mining, manufacturing, wholesale trade, and certain retail industries in the fifty U.S. states and the District of Columbia. The budget for the 2002 CFS survey cycle is estimated at \$13 million. It is by far the most comprehensive freight flow database in the U.S., providing freight data in terms of tons, ton-miles, value, shipment distance, commodity, and weight for all major transportation modes (i.e., air, truck, rail, water, and pipeline), as well as intermodal combinations. Unfortunately, budget constraints have resulted in the sample size of the CFS survey being reduced from 200,000 (1993) to 100,000 (1997), and 50,000 (2002) —the latter represents approximately 7 percent of the 750,000 industries registered in the Census Bureau's Business Register and that are covered by the CFS. In addition, concerns exist about the adequate coverage of agricultural shipments (i.e., the movement from the farm to the first point of assembly), shipments from transportation and service businesses, and most of the retail businesses (National Research Council of the National Academies, 2003). Finally, the CFS database provides limited information on foreign shipments, because only domestic shippers are sampled. For example, imports are captured only when transported from the importer's location to a destination elsewhere in the U.S. Thus, the movement between the port of entry and the importer's location is not captured in the CFS database. Similarly for exports, the export destination is provided as the port of exit in the U.S. (National Research Council of the National Academies, 2003).

3.2.3 Robust Regression Analysis Approach

The literature revealed that regression analysis has been employed to generate disaggregated county freight flows from the publicly available CFS database. The approach is essentially identical to the trip generation and distribution steps in the four-step travel demand modeling process. The first step is to develop regression equations for freight trip generations and attractions as functions of a set of explanatory variables that impact commodity productions and attractions in a county. County employment and population data are the commonly used explanatory variables for developing regression equations for freight generations and attractions. Since the public freight data are not accurate owing to sampling errors and data disclosure issues, an ordinary least squares (OLS) regression will probably provide inaccurate estimates of the regression coefficients owing to the presence of outliers. To account for the presence of outliers, robust regression analysis has been recommended. Robust regression is an iterative regression process that identifies outliers in the data and minimizes their effects in the estimation of the

regression coefficients. The outliers are excluded from the model, and weighted stepwise regression is performed to estimate the final regression coefficients.

Freight Generation Equations

Robust freight generation equations¹ are developed with the objective of predicting total commodity productions in a state as a function of the production employment. The CFS database provides detailed freight flows (i.e., annual commodity tonnage) from production state i to attraction states j , P_{ij}^k , $j = 1$ to n . Total productions of commodity k in state i , P_i^k , can thus be calculated as follows:

$$P_i^k = \sum_{j=1}^n P_{ij}^k$$

where n = number of attraction states.

The dependent variable in the regression equations is the total state productions of commodity k derived from the CFS database, which is modeled as a function of the state employment data associated with the production of commodity k . The state employment data associated with the “commodity productions” can be obtained from the Occupational Employment Statistics (OES) published by the Bureau of Labor Statistics. It is available online at: <http://www.bls.gov/oes/current/oesrcst.htm>. Freight generation equations have to be developed for each commodity category k . The data set for the regression analysis consists of total productions in each state and the corresponding employment associated with the production of commodity k . Robust regression is subsequently performed to estimate the parameters for the freight generation equations.

Freight Attraction Equations

Freight attraction equations² are developed with the objective to predict the total commodity attractions to a state as a function of, for example, population. The CFS database provides detailed freight flows (i.e., annual commodity tonnage) to attraction state j from production states i , A_{ji}^k , $i = 1$ to n . Total attractions of commodity k to state j can thus be calculated as follows:

$$A_j^k = \sum_{i=1}^n A_{ji}^k$$

where n = number of production states.

Freight attraction equations are thus developed by performing robust regressions of total commodity attractions with the state population as the explanatory variable for each

¹ Freight generation equations for commodity productions cannot be developed at the county level, because the CFS does not capture county commodity flows. The freight generation equations are thus developed at the state level as a function of state employment.

² Freight attraction equations for commodity attractions cannot be developed at the county level, because the CFS does not capture county commodity flows. The freight attraction equations are thus developed at the state level as a function of state population.

commodity category. Population data are available online from the state population data sets (<http://www.census.gov/popest/states/NST-ann-est.html>) developed by the U.S. Bureau of the Census as part of its Population Estimates Program.

Freight Distribution

The estimated freight generation and attraction equations can be used to estimate the total county productions and attractions for each commodity category k , by using the county employment data collected from the County Business Patterns and the county population statistics from the Population Estimates Program of the Census Bureau as inputs.

The gravity model, commonly used for freight flow distribution, is founded in the theory that interzonal freight flows are directly proportional to the production/attraction levels of origin/destination zones and inversely proportional to a measure of impedance. Travel time or transportation cost is considered a good measure of impedance because freight flows between zones are impacted by both. The mathematical equation for the gravity model is as follows:

$$T_{ij}^k = P_i^k * \frac{A_j^k F_{ij}}{\sum_{j=1}^n A_j^k F_{ij}}$$

where

T_{ij}^k = total flows of commodity k from zone i to zone j ,

P_i^k = total productions of commodity k in zone i ,

A_j^k = total attractions of commodity k in zone j ,

F_{ij} = impedance factor, and

n = number of zones.

The Mississippi Department of Transportation used the CFS data and a gravity model to estimate county freight flows. The approach is summarized in Appendix B.

3.3 Concluding Remarks

The gravity model assumes the distribution of truck flows is a function of the production/attraction levels at the origins and destinations and an impedance measure. The model therefore does not consider the influence of other factors, such as long standing trade relationships, shipper logistics patterns, or the affect of large freight facilities, such as warehouses, distribution centers, and intermodal terminals, on the distribution of freight flows. Also, calibration of the gravity model requires arriving at the parameters of the gamma function of the impedance measure through successive iterations. Parameter estimation can thus be cumbersome, involving successive iterations until the calculated flows converge to the observed values. The next chapter explores the use of a multinomial logit model (MNL) to estimate disaggregated county-level truck travel data for Texas from the CFS database.

4. Proposed Methodology for Populating Statewide Analysis Model Database Using Commodity Flow Survey Data

Given available truck data from the Commodity Flow Survey (CFS) and county economic data from the IMPLAN databases, multinomial logit (MNL) models were used to estimate county-level truck travel data for Texas. This chapter provides an overview of the MNL models and discusses the steps involved in estimating disaggregated truck data for Texas using the MNL model approach. In addition, this chapter documents the strengths and limitations of the modeling approach and concludes with a comparison of the model output with the truck data contained in the Reebe TRANSEARCH database.

4.1 Multinomial Logit Model

The logit model is a binary choice model (i.e., two alternatives) and assumes that the error terms of the utility functions³ are independent and Gumbel⁴ distributed. The probability of choosing a particular alternative (a_1) in lieu of the other (a_2) is given by Equation 1:

$$P(a_1) = \frac{e^{\mu V_1}}{e^{\mu V_1} + e^{\mu V_2}} \quad (1)$$

where

μ is the scale parameter⁵ of the error term of the utility functions (commonly assumed as 1),

V_1 is the deterministic utility associated with alternative 1, and

V_2 is the deterministic utility associated with alternative 2.

The deterministic utility of an alternative is a function of the attributes associated with that alternative and of the decision-maker and is thus a measure of the propensity of the decision-maker to choose the alternative. The deterministic utility is generally assumed to be a linear function of the attributes for simplicity in formulation and estimation of the model (linear-in-parameters logit model). Truck flows between origins and destinations can thus be modeled using the deterministic utility, which is a function of the origin and destination attributes that impact freight movements. In the case of n attributes affecting the choice of the decision-maker, the deterministic utility is given by Equation 2:

³ The utility function presents a measure of individual preference for a specific alternative. In other words, if an individual has to choose among “n” alternatives, he/she will choose the alternative that maximizes his/her utility function.

⁴ The MNL model assumes that the error terms of the utility function follow a Gumbel distribution – a type of probability distribution for random variables. This allows for the probability of the individual to choose a particular alternative to be expressed by Equation 1.

⁵ Probability distributions of random variables have two parameters: a location and a scale parameter. The scale parameter describes the “shape” of the probability distribution (i.e., more spread or compressed). The standard form of the probability distribution has a location parameter equal to zero and a scale parameter equal to one. In calibrating the MNL, the standard form of the Gumbel distribution is used therefore, the value of the scale parameter is equal to one.

$$V_a = \sum_{i=1}^n \beta_i X_{a_i} \quad (2)$$

The multinomial logit model is a generalization of the logit model with more than two alternatives. Extending the binary choice to a case with m alternatives, where V_i is the deterministic utility associated with alternative i , the probability of choosing alternative i among m alternatives is given by Equation 3:

$$P(a_i) = \frac{e^{\mu V_i}}{\sum_{k=1}^m e^{\mu V_k}} \quad (3)$$

where V_i is the deterministic utility associated with alternative i .

4.1.1 Multinomial Logit Model for Truck Flow Distribution

The MNL model approach is particularly suitable for modeling the distribution of truck flows on the highway network. Let us consider the case of n zones with commodity productions and attractions, having interzonal truck freight movements. The distribution of truck flows among these zones needs to be modeled for the following cases:

- Commodity productions (i.e., production flow distribution): Truck flows from production zone i to n attraction zones.
- Commodity attractions (i.e., attraction flow distribution): Truck flows to attraction zone j from n production zones.

Commodity Productions

The production flow distribution of commodities can be modeled as a function of the generalized cost of transportation and the relative attraction level of the destination zones. Similarly, the attraction flow distribution of commodities can be modeled as a function of the generalized cost of transportation and the relative production level of the origin zones. Owing to a lack of generalized cost data, centroidal distances between zones were employed as the impedance measure affecting freight flow distribution.

The MNL model is developed separately to model the distribution of freight flows from production zone i to the n attraction zones and freight flows to attraction zone j from the n production zones for each commodity group k .

For truck flows from production zone i to the n attraction zones for commodity group k , a deterministic utility function is defined for each of the n alternatives (attraction zones) representing the propensity for commodity flows to each attraction zone. Considering the flows from zone i to attraction zone j for commodity group k , the deterministic utility is defined by Equation 4:

$$V_{ij}^k = \alpha_{0j}^k + \alpha_1^k d_{ij} + \alpha_2^k (FA_j^k) \quad (4)$$

where

α_{0j} , α_1 , and α_2 are the coefficients to be determined during model calibration,

d_{ij} is the centroidal distance between zones i and j , and

FA_j^k is the relative attraction level of zone j for commodity group k .

$$FA_j^k = \frac{A_j^k}{\sum_{j=1}^n A_j^k} \quad (5)$$

where A_j^k = total attractions of commodity k in zone j .

FA_j^k thus represents the attraction of commodity group k in zone j relative to the total attractions of commodity group k in all the attraction zones.

A constant term (α_0) is included in the linear utility function to account for all the factors that may affect freight flows from the production zone to the attraction zones that have not been included in the model.

Given the linear utility function (V_{ij}^k) and assuming the scale parameter of the error terms of the MNL model is unity ($\mu = 1$), the distribution of total productions of commodity group k in zone i , P_i^k , to each of the attraction zones can be computed. For example, total truck flows of commodity group k from zone i to zone j , T_{ij}^k , is given by Equation 6:

$$T_{ij}^k = P_i^k * \frac{e^{V_{ij}^k}}{\sum_{j=1}^n e^{V_{ij}^k}} \quad (6)$$

where P_i^k = total productions of commodity k in zone i .

Commodity Attractions

For the attraction flow distribution model, the deterministic utilities are defined for truck flows of commodity k to attraction zone j from each production zone i by the following linear function:

$$V_{ji}^k = \beta_{0i}^k + \beta_1^k d_{ji} + \beta_2^k (FP_i^k) \quad (7)$$

where

β_{0i} , β_1 and β_2 are the coefficients to be determined during model calibration,

d_{ji} is the centroidal distance between zones j and i , and

FP_i^k is the relative production level of commodity group k in zone i .

$$FP_i^k = \frac{P_i^k}{\sum_{i=1}^n P_i^k} \quad (8)$$

where FP_i^k represents the total productions of commodity group k in zone i relative to the total productions of commodity group k in all the production zones.

A constant term (β_0) has been included in the linear utility function to account for all the factors that may be affecting the attraction flow distribution of trucks that have not been included in the model.

Given the linear utility function (V_{ji}^k) and assuming the scale parameter of the error terms of the MNL model is unity ($\mu = 1$), the distribution of total attractions of commodity group k to zone j , A_j^k , from each of the production zones can be computed. For example, truck flows of commodity group k to zone j from production zone i , T_{ji}^k , is given by Equation 9:

$$T_{ji}^k = A_j^k * \frac{e^{V_{ji}^k}}{\sum_{i=1}^n e^{V_{ji}^k}} \quad (9)$$

4.2 Calibration of Multinomial Logit Model to Estimate Truck Flows for Texas

The multinomial logit model needs to be calibrated at the state level before it can be used for truck flow disaggregation at the county level. State-level MNL models were developed for the following truck flow distribution cases for each commodity category:

- Production flow distribution. Distribution of truck flows from production state to the n attraction states.
- Attraction flow distribution. Distribution of truck flows to attraction state from the n production states.

4.2.1 Production Flow Distribution Model

The MNL production flow distribution model estimates the fraction of the total productions in a state moving to each attraction state by truck on the basis of the attributes of the attraction state and the generalized cost of transportation. The centroidal distance between states along the highway network and a measure of the attraction level of the destination states were considered to be important factors in determining truck flows from the production state to the n attraction states.

Data Requirements for Model Calibration

The following data were used for calibrating the production flow distribution model:

- 1997 Commodity Flow Survey (CFS)
- Annual tonnage of truck flows from production state i ($i = 1-50$) to attraction state j ($j = 1-50$) for each commodity category k (T_{ij}^k)
- Relative attraction level of state j ($j = 1-50$) for each commodity category k (FA_j^k)

$$FA_j^k = \frac{\sum_{i=1}^{50} T_{ij}^k}{\sum_{j=1}^{50} \sum_{i=1}^{50} T_{ij}^k} \quad (10)$$

- Interstate centroidal distances computed from TransCAD GIS software
 - The interstate centroidal distances (d_{ij}) were computed using the TransCAD software as the shortest path distance along the highway network between state centroids.

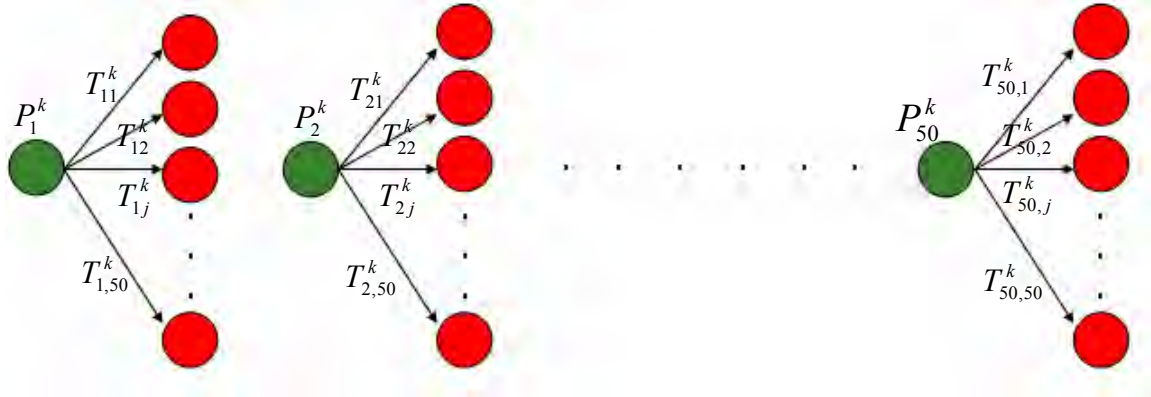
Interzonal Centroid Distance Computations

The interzonal centroidal distances (d_{ij}) are computed, using the TransCAD software, as the shortest path distance along the highway network between zone centroids. Because the existing highway network is not linked to the zone centroids, centroids are connected to the highway layer using *centroid connectors* in the *Planning/Planning Utilities* feature of TransCAD. After including the centroids in the highway node layer, a new highway network is created by including the centroid nodes in the highway line layer. This is done by selecting the highway nodes representing the centroids and designating these nodes as centroids in the *Network Settings* feature in TransCAD. Once the highway network is created with the state centroids, the interstate centroidal distances are computed as a multiple shortest path matrix by minimizing the highway network distance between the centroids.

Production Flow Distribution Model Calibration Methodology

The calibration of the MNL production flow distribution model involves computing the coefficients of the utility functions for the attraction states using the state-to-state CFS truck flow data for each commodity category k .

The truck flows from each production state i ($i = 1-50$) to the fifty attraction states for commodity k can be represented as follows:



$$P_i^k = \sum_{j=1}^{50} T_{ij}^k \quad (11)$$

where

P_i^k = total productions of commodity k in state i ($i = 1-50$), and

T_{ij}^k = truck flows of commodity k from production state i to attraction state j ($i, j = 1-50$).

In the MNL production flow distribution model, the deterministic utility for truck flows of commodity k from production state i to attraction state j ($j = 1-50$) is given by Equation 12:

$$V_{ij}^k = \alpha_{0j}^k + \alpha_1^k d_{ij} + \alpha_2^k FA_j^k \quad (12)$$

where

α_{0j}^k , α_1^k and α_2^k are the coefficients to be determined during model calibration,

d_{ij} is the centroidal distance between states i and j , and

FA_j^k is the relative attraction level of state j for commodity k .

The fraction of the total productions of commodity k in state i flowing to state j is thus given by Equation 13:

$$T_{ij}^k = P_i^k * \frac{e^{V_{ij}^k}}{\sum_{j=1}^{50} e^{V_{ij}^k}} \quad (13)$$

The values of T_{ij}^k and P_i^k in terms of annual tonnage are available from the CFS. The only unknowns in Equation 13 are the deterministic utility values (V_{ij}^k). Because $\sum_{j=1}^{50} T_{ij}^k = P_i^k$ by flow equilibrium, the MNL production flow distribution model comprises forty-nine degrees of freedom (redundancy = 1) with fifty unknown deterministic utilities. To determine the utility values for truck flows from state i to state j ($j = 1-50$), the utility for truck flows from state i to

an arbitrary state s ($s \in j$) is assigned the value zero ($V_{is}^k = 0$). The utilities for truck flows to the remaining states j ($j = 1-50, j \neq s$) are then computed relative to the utility for truck flows from state i to state s .

Substituting the utility for truck flows of commodity k from state i to state s (V_{is}^k) in Equation 13 results in Equation 14:

$$T_{is}^k = P_i^k * \frac{e^0}{\sum_{j=1}^{50} e^{V_{ij}^k}} = P_i^k * \frac{1}{\sum_{j=1}^{50} e^{V_{ij}^k}} \quad (14)$$

where

T_{is}^k = truck flows of commodity k from state i to state s , and

V_{ij}^k = utility for truck flows from state i to attraction state j relative to attraction state s .

From Equation 14, it follows that

$$\sum_{j=1}^{50} e^{V_{ij}^k} = \frac{P_i^k}{T_{is}^k} \quad (15)$$

Substituting for $\sum_{j=1}^{50} e^{V_{ij}^k}$ in Equation 13 and taking the natural logarithm results in Equation 16:

$$V_{ij}^k = \ln\left(\frac{T_{ij}^k}{T_{is}^k}\right), i = 1 \text{ to } 50, j = 1 \text{ to } 50, j \neq s \quad (16)$$

Using the CFS data for T_{ij}^k ($i = 1-50, j = 1-50, j \neq s$) and T_{is}^k , the deterministic utilities for truck flows of commodity k from state i to state j ($j \neq s$; V_{ij}^k) can be calculated using Equation 16. The utility values indicate the propensity of states j ($j \neq s$) to attract flows from production state i , relative to attraction state s . The relative utilities can now be expressed as a linear function of the explanatory variables, as follows:

$$V_{ij}^k = \alpha_{0j}^k + \alpha_1^k d_{ij|s} + \alpha_2^k FA_{j|s}^k \quad (17)$$

where

V_{ij}^k = deterministic utility for truck flows of commodity k from state i to state j , relative to truck flows to state s ,

$d_{ij|s} = d_{ij} - d_{is}$ = relative centroidal distance of state i between states j and s , and

$FA_{j|s}^k = FA_j^k - FA_s^k$ = total attraction level of state j relative to state s for commodity k .

The coefficients of the utility functions for each commodity k were estimated using a linear regression of the dependent variable V_{ij}^k with the independent variables $d_{ij|s}$ and $FA_{j|s}^k$.

4.2.2 Attraction Flow Distribution Model

The MNL attraction flow distribution model estimates the fraction of the total attractions in state j ($j = 1-50$) originating from the fifty production states by truck for each commodity category k . The interstate centroidal distances and the relative production levels of the origin states were considered as the primary explanatory variables impacting the distribution of truck flows to the attraction state j ($j = 1, 2, \dots, 50$) from the fifty production states.

Data Requirements for Model Calibration

The following data were used for calibrating the attraction flow distribution model:

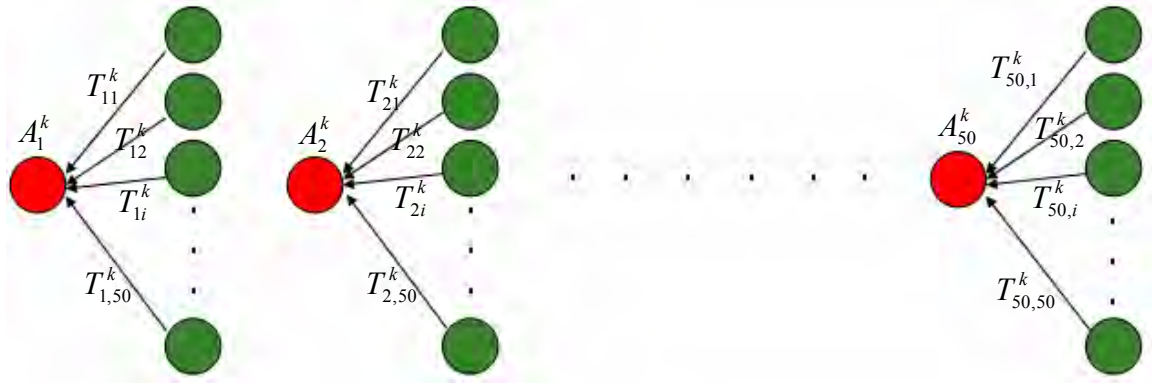
- 1997 CFS
 - Annual truck tonnage to attraction state j ($j = 1-50$) from the fifty production states for each commodity category k (T_{ji}^k)
 - Relative production level of state i ($i = 1-50$) for each commodity category k (FP_i^k)

$$FP_i^k = \frac{\sum_{j=1}^{50} T_{ji}^k}{\sum_{i=1}^{50} \sum_{j=1}^{50} T_{ji}^k} \quad (18)$$

- Interstate centroidal distances derived from TransCAD GIS software.

Attraction Flow Distribution Model Calibration Methodology

Calibration of the MNL attraction flow distribution model involves computing the coefficients of the utility functions for the production states using the CFS truck tonnage data for truck flows to the attraction states j ($j = 1-50$) from the fifty production states for each commodity category k . The annual truck tonnage flow distribution attracted to state j ($j = 1-50$) from the fifty production states for commodity k can be illustrated as follows:



$$A_j^k = \sum_{i=1}^{50} T_{ji}^k \quad (19)$$

where

A_j^k = total truck attractions (tons) of commodity k to state j ($j = 1-50$), and

T_{ji}^k = annual truck tonnage of commodity k attracted to state j from production state i .

From the attraction flow distribution model, the fraction of the total attractions of commodity k to state j from production state i ($i = 1-50$) is given by Equation 20:

$$T_{ji}^k = A_j^k * \frac{e^{V_{ji}^k}}{\sum_{i=1}^{50} e^{V_{ji}^k}} \quad (20)$$

where

V_{ji}^k = deterministic utility for truck flows of commodity k to state j from state i ($i = 1-50$), which is in turn given by Equation 21:

$$V_{ji}^k = \beta_{0i}^k + \beta_1^k d_{ji} + \beta_2^k FP_i^k \quad (21)$$

where

β_{0i}^k, β_1^k and β_2^k are the coefficients to be determined during model calibration,

d_{ji} is the centroidal distance between states j and i , and

FP_i^k is the relative production level of states i of commodity group k .

The fraction of the total attractions of commodity k to state j from production state i ($i = 1-50$) is thus given by Equation 22:

$$T_{ji}^k = A_j^k * \frac{e^{V_{ji}^k}}{\sum_{i=1}^{50} e^{V_{ji}^k}} \quad (22)$$

The values for T_{ji}^k and A_j^k can be derived from the CFS. The only unknowns in the above equation are thus the deterministic utilities V_{ji}^k . As $\sum_{i=1}^{50} T_{ji}^k = A_j^k$ by flow equilibrium, the MNL model has forty-nine degrees of freedom (redundancy = 1) with fifty unknown deterministic utilities. To determine the utility values for truck flows to state j from production states i ($i = 1-50$), the utility for truck flows to state j from an arbitrary production state r ($r \in j$) is assigned the value zero ($V_{jr}^k = 0$). The utilities for truck flows to state j from the remaining states i ($i = 1-50, i \neq r$) are then computed relative to the utility for truck flows to state j from state r . Substituting the utility for truck flows of commodity k to state j from production state r in Equation 22 results in Equation 23:

$$T_{jr}^k = A_j^k * \frac{e^0}{\sum_{i=1}^{50} e^{V_{ji}^k}} = A_j^k * \frac{1}{\sum_{i=1}^{50} e^{V_{ji}^k}} \quad (23)$$

where

T_{jr}^k = truck flows of commodity k to state j from production state r , and

V_{ji}^k = deterministic utility for truck flows of commodity k to state j from state i relative to flows from state r .

From Equation 23, it follows that

$$\sum_{i=1}^{50} e^{V_{ji}^k} = \frac{A_j^k}{T_{jr}^k} \quad (24)$$

Substituting $\sum_{i=1}^{50} e^{V_{ji}^k}$ in Equation 22 and taking the natural logarithm gives Equation 25:

$$V_{ji}^k = \ln\left(\frac{T_{ji}^k}{T_{jr}^k}\right), j = 1 \text{ to } 50, i = 1 \text{ to } 50, i \neq r \quad (25)$$

Using the CFS data for T_{ji}^k ($j = 1-50, i = 1-50, i \neq r$) and T_{jr}^k ($j = 1-50$), the relative utilities for truck flows of commodity k to state j from production state i ($i = 1-50; V_{ji}^k$) can be calculated using Equation 25. The deterministic utilities can now be expressed as a linear function of the explanatory variables as follows:

$$V_{ji}^k = \beta_{0i}^k + \beta_1^k d_{ji|r} + \beta_2^k FP_{i|r}^k \quad (26)$$

where

V_{ji}^k = deterministic utility for truck flows of commodity k to state j from state i , relative to flows from state r ,

$d_{ji|r} = d_{ji} - d_{jr}$ = relative centroidal distance of state j between states i and r , and

$FP_{i|r}^k = FP_i^k - FP_r^k$ = total production level of state i relative to state r of commodity k .

The coefficients of the utility functions for each commodity k can be estimated using linear regression of the dependent variable V_{ji}^k with the independent variables $d_{ji|r}$ and $FP_{i|r}^k$.

4.3 Truck Flow Disaggregation to the County Level

The objective of the state-level model calibration is to generate county-level truck flows for Texas from the available CFS truck flow data. The calibrated state-level MNL production and attraction flow distribution models are thus used to estimate the internal–external, external–internal, and internal–internal truck flows in Texas at the county level. In other words, the following truck flows are estimated during the truck flow disaggregation step:

- Texas county-to-state truck flows of commodity k
- State-to-Texas county truck flows of commodity k
- Texas county-to-county truck flows of commodity k .

4.3.1 Data Requirements for Estimating County-Level Truck Flow Data

The following data are required for disaggregating truck flows to the Texas county level.

- 1997 CFS truck flow data
 - Texas-to-state truck flows for each commodity category
 - State-to-Texas truck flows for each commodity category
 - Texas-to-Texas truck flows for each commodity category
- IMPLAN⁶ economic impact modeling system. The following county-level economic data were used from the 1998 IMPLAN databases published by the Minnesota IMPLAN Group, Inc. (MIG, Inc):
 - Texas county productions by commodity category (millions of dollars) derived from the Output field in IMPLAN
 - Texas county attractions by commodity category (millions of dollars) derived from the institutional commodity demand field in IMPLAN

⁶ The IMPLAN data were used to determine what fraction of each of the commodities was produced in or attracted to each Texas county. This fraction is one of the variables in the deterministic utility function (see Equation 29) that is used to disaggregate the state-to-state CFS commodity flows to the county level. The IMPLAN data were used because it was the only readily available source from which Texas county commodity productions and attractions could be derived. For additional information about the IMPLAN data, the reader is referred to the following website: <http://www.implan.com/what/html>.

- Impedance measure
 - County-to-state centroidal distances (miles)
 - County-to-county centroidal distances (miles)

4.3.2 Texas County-to-State Flows

Texas county-to-state flows are estimated from the Texas-to-state truck flows by commodity category captured in the CFS and by applying the calibrated MNL attraction flow distribution model. The fraction of the total truck flows of commodity k to state j from Texas, originating from county i ($i = 1-254$) is a function of the centroidal distance between county i and state j , and the relative production level of county i of commodity k . The utility function for truck flows to attraction state j from Texas (V_{jT}^k) is derived from the calibrated attraction flow distribution model:

$$V_{jT}^k = \beta_{0T}^k + \beta_1^k d_{jT} + \beta_2^k FP_T^k \quad (27)$$

where

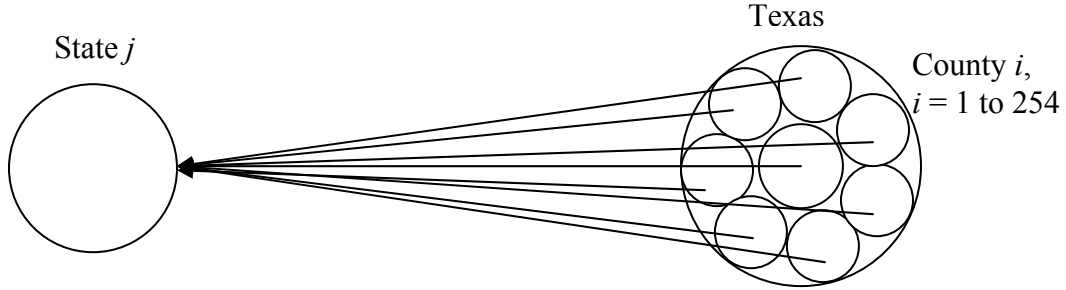
V_{jT}^k = deterministic utility for flows to state j from Texas for commodity k ,

β_{0T}^k , β_1^k and β_2^k are the utility coefficients derived from model calibration,

d_{jT} = centroidal distance between state j and Texas, and

FP_T^k = relative production level in Texas of commodity k .

The truck flow distribution to attraction state j from Texas counties can be illustrated as follows:



By applying the MNL attraction flow distribution model, the fraction of the total attractions of commodity k in state j from Texas, originating from each county i ($i = 1-254$; $T_{jC_i}^k$) can be determined by Equation 28:

$$T_{jC_i}^k = T_{jT}^k * \frac{e^{V_{jC_i}^k}}{\sum_{i=1}^{254} e^{V_{jC_i}^k}} \quad (28)$$

where

$T_{jC_i}^k$ = total truck flows of commodity k to state j from county i in Texas,

T_{jT}^k = total truck flows of commodity k to state j from Texas, and

$V_{jC_i}^k$ = deterministic utility for truck flows of commodity k to state j from county i in Texas, which in turn is given by Equation 29:

$$V_{jC_i}^k = \beta_{0T}^k + \beta_1^k d_{jC_i} + \beta_2^k FP_{C_i}^k \quad (29)$$

where

d_{jC_i} = centroidal distance between state j and county i in Texas, and

$FP_{C_i}^k$ = relative production level of county i for commodity k from IMPLAN database, which in turn is given by Equation 30:

$$FP_{C_i}^k = \frac{P_{C_i}^k}{\sum_{i=1}^{254} P_{C_i}^k} \quad (30)$$

By substituting the deterministic utility functions in the model, the annual truck tonnage attracted to state j from county i for commodity k can be estimated from Equation 31:

$$T_{jC_i}^k = T_{jT}^k * \frac{e^{\beta_{0T}^k + \beta_1^k d_{jC_i} + \beta_2^k FP_{C_i}^k}}{\sum_{i=1}^{254} e^{\beta_{0T}^k + \beta_1^k d_{jC_i} + \beta_2^k FP_{C_i}^k}} \quad (31)$$

4.3.3 State-to-Texas County Flows

State-to-Texas county truck flows are estimated from the state-to-Texas truck flows by commodity category captured in the CFS by applying the calibrated MNL production distribution flow model. In other words, out of the total truck flows from production state i ($i = 1-49$, $i \neq \text{Texas}$) to Texas, the fraction of the flows destined for each Texas county needs to be determined.

The fraction of the total truck flows from origin state i to Texas, destined to county j ($j = 1-254$) is a function of the centroidal distance between state i and county j , and the relative attraction level in county j for commodity k . The utility function for truck flows of commodity k from production state i to Texas (V_{iT}^k) is derived from the calibrated production flow distribution model:

$$V_{iT}^k = \alpha_{0T}^k + \alpha_1^k d_{iT} + \alpha_2^k FA_T^k \quad (32)$$

where

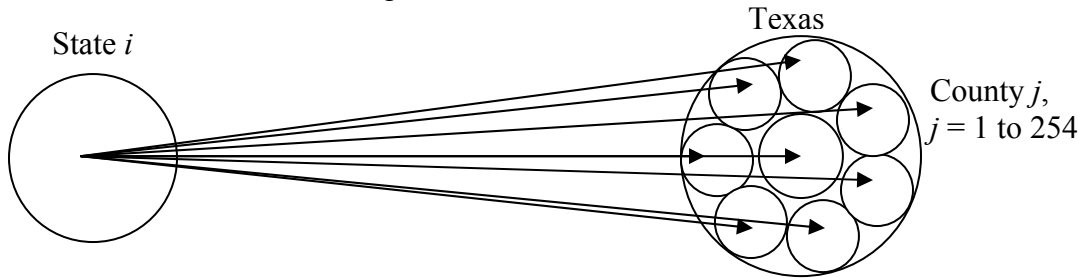
V_{iT}^k = deterministic utility for truck flows from production state i to Texas for commodity k ,

α_{0T}^k , α_1^k and α_2^k = utility coefficients derived from model calibration,

d_{iT} = centroidal distance between state i and Texas, and

FA_T^k = relative attraction level of Texas for commodity k .

The truck flow distribution from production state i to Texas can be illustrated as follows:



By applying the MNL production flow distribution model, the fraction of the total truck flows from state i to Texas, destined for each Texas county j ($j = 1-254$) for commodity k can be determined by Equation 33:

$$T_{iC_j}^k = T_{iT}^k * \frac{e^{V_{iC_j}^k}}{\sum_{j=1}^{254} e^{V_{iC_j}^k}} \quad (33)$$

where

$T_{iC_j}^k$ = total truck flows of commodity k from production state i to Texas county j ,

T_{iT}^k = total truck flows of commodity k from state i to Texas, and

$V_{iC_j}^k$ = deterministic utility function for truck flows of commodity k from production state i to Texas county j , which in turn is given by Equation 34:

$$V_{iC_j}^k = \alpha_{0T}^k + \alpha_1^k d_{iC_j} + \alpha_2^k FA_{C_j}^k \quad (34)$$

where

d_{iC_j} = centroidal distance between state i and county j in Texas, and

$FA_{C_j}^k$ = relative attraction level of county j for commodity k from IMPLAN database.

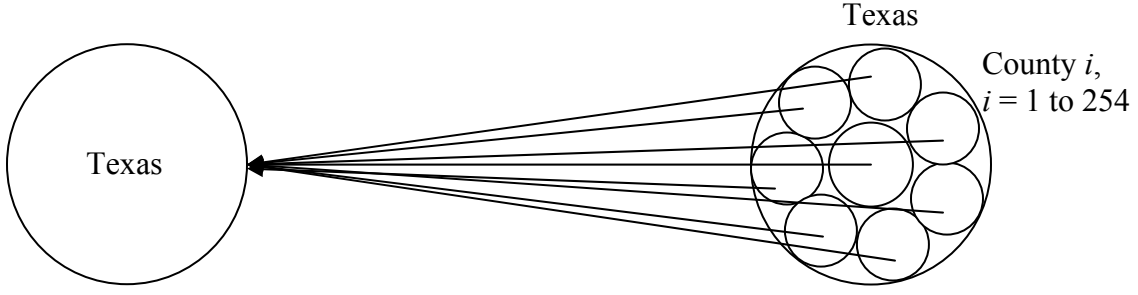
$$FA_{C_j}^k = \frac{A_{C_j}^k}{\sum_{j=1}^{254} A_{C_j}^k} \quad (35)$$

By substituting the deterministic utility functions in the model, the annual truck tonnage from production state i to Texas county j for commodity k can be estimated from Equation 36:

$$T_{iC_j}^k = T_{iT}^k * \frac{e^{\alpha_{0T}^k + \alpha_1^k d_{iC_j} + \alpha_2^k FA_{C_j}^k}}{\sum_{j=1}^{254} e^{\alpha_{0T}^k + \alpha_1^k d_{iC_j} + \alpha_2^k FA_{C_j}^k}} \quad (36)$$

4.3.4 Texas County-to-County Flows

Texas county-to-county truck flows were estimated from the Texas-to-Texas truck flow data by commodity category captured in the CFS and the calibrated MNL attraction and production flow distribution models. County-to-county truck flows were thus generated following a two-step procedure. First, the fractions of the total intrastate Texas attractions originating from each county i ($i = 1-254$) are determined by applying the attraction flow distribution model. The distribution of the intrastate Texas attractions originating from each county can be illustrated as follows:



Applying the MNL attraction flow distribution model, intrastate truck flows attracted to Texas from each county i ($i = 1-254$) can be calculated as follows:

$$T_{TC_i}^k = T_{TT}^k * \frac{e^{V_{TC_i}^k}}{\sum_{i=1}^{254} e^{V_{TC_i}^k}} \quad (37)$$

where

$T_{TC_i}^k$ = truck flows attracted to Texas from county i for commodity k ,

T_{TT}^k = total Texas-to-Texas (intrastate) truck flows for commodity k , and

$V_{TC_i}^k$ = deterministic utility for intrastate truck flows from county i for commodity k , which in turn is given by Equation 38:

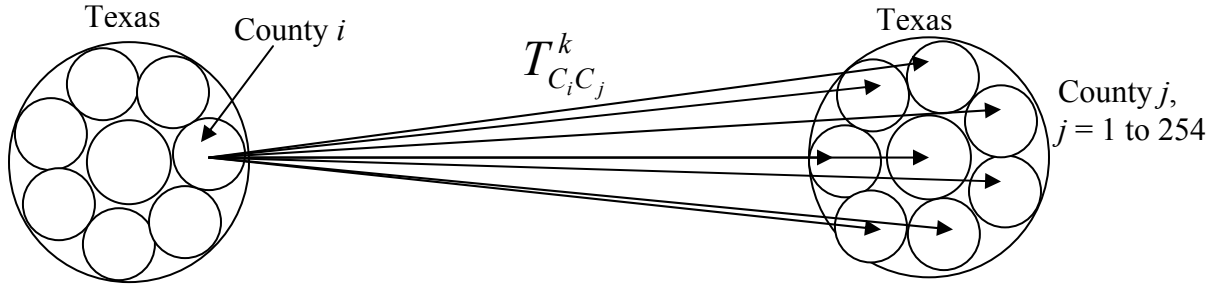
$$V_{TC_i}^k = \beta_{0T}^k + \beta_1^k d_{TC_i} + \beta_2^k FP_{C_i}^k \quad (38)$$

where

d_{TC_i} = centroidal distance between Texas and county i , and

$FP_{C_i}^k$ = relative production level of county i for commodity k .

Second, county-to-county truck flows are estimated by distributing the intrastate truck flows originating in each county i ($i = 1-254$) to each destination county j ($j = 1-254$) using the production flow distribution model. The distribution of intrastate truck flows originating in each county i ($i = 1-254$) to each destination county j ($j = 1-254$) can be presented as follows:



Applying the MNL production flow distribution model, truck flows of commodity k from county i to county j can be computed as follows:

$$T_{C_i C_j}^k = T_{C_i T}^k * \frac{e^{V_{C_i C_j}^k}}{\sum_{j=1}^{254} e^{V_{C_i C_j}^k}} \quad (39)$$

where

$T_{C_i C_j}^k$ = truck flows of commodity k from county i to county j ,

$T_{C_i T}^k$ = truck flows of commodity k from county i to Texas calculated in Step 1, and

$V_{C_i C_j}^k$ = deterministic utility for truck flows of commodity k from county i to county j .

$$V_{C_i C_j}^k = \alpha_{0T}^k + \alpha_1^k d_{C_i C_j} + \alpha_2^k FA_{C_j}^k \quad (40)$$

where

$d_{C_i C_j}$ = centroidal distance between county i and county j .

$FA_{C_j}^k$ = relative attraction level of county j for commodity k from IMPLAN database.

By substituting the deterministic utility function in the model, the truck flows (tonnage) from county i to county j for commodity k can be estimated from Equation 41.

$$T_{C_i C_j}^k = T_{C_i T}^k * \frac{e^{\alpha_{0T}^k + \alpha_1^k d_{C_i C_j} + \alpha_2^k FA_{C_j}^k}}{\sum_{j=1}^{254} e^{\alpha_{0T}^k + \alpha_1^k d_{C_i C_j} + \alpha_2^k FA_{C_j}^k}} \quad (41)$$

4.4 Results Of Model Calibration

4.4.1 MNL Production Flow Distribution Model

The calibration of the MNL production flow distribution model provides the coefficients of the linear utility functions for truck flows from production state i to each of the fifty attraction states. MNL models were calibrated for the following commodity categories.

- Agriculture
- Food
- Building Materials
- Raw Materials
- Chemicals and Petroleum
- Wood
- Textiles
- Machinery
- Miscellaneous Commodities

Production and attraction flow distribution models were not developed for hazardous materials (HAZMAT) and secondary shipments, owing to inadequate truck flow data.

The linear utility function representing the propensity of truck flows of commodity k from production state i to attraction state j (V_{ij}^k) among fifty alternatives is given by Equation 42:

$$V_{ij}^k = \alpha_{0j}^k + \alpha_1^k d_{ij} + \alpha_2^k FA_j^k \quad (42)$$

where

d_{ij} is the centroidal distance between states i and j , and

FA_j^k is the relative attraction level of state j for commodity k .

The coefficients of the level of service variables (i.e., centroidal distance between state i and state j [d_{ij}] and the attraction level of state j for commodity k [FA_j^k]) in the linear utility function are common to all alternatives. The constant term is, however, specific to each attraction state because it accounts for factors other than distance and relative attraction level that may be different for each attraction state.

The linear utility function for truck flows of commodity k from production state i to attraction state j can be expressed as Equation 43:

$$V_{ij}^k = \alpha_{01}^k X_{01} + \alpha_{02}^k X_{02} + \dots + \alpha_{050}^k X_{50} + \alpha_1^k d_{ij} + \alpha_2^k FA_j^k \quad (43)$$

where

X_{0n} , $n = 1, 2, \dots, 50$ denote dummy variables with values:

$$X_{0n} = 1 \text{ if } j = n$$

$$X_{0n} = 0 \text{ if } j \neq n$$

The dependent variables are expressed in terms of the relative utilities as calculated from the CFS data on truck flow distribution from the production states. The coefficients of the linear utility functions for the production flow distribution model can thus be estimated by performing linear regression of V_{ij}^k with X_{0n} , $n = 1, 2, \dots, 50$, $d_{ij|s}$ and $FA_{j|s}^k$ — $d_{ij|s}$ and $FA_{j|s}^k$ as defined during the model calibration phase. The OLS linear regression model for the utility functions is given by Equation 44:

$$V_{ij}^k = \hat{\alpha}_{01}^k X_{01} + \hat{\alpha}_{02}^k X_{02} + \dots + \hat{\alpha}_{050}^k X_{050} + \hat{\alpha}_1^k d_{ij|s} + \hat{\alpha}_2^k FA_{j|s}^k \quad (44)$$

where

V_{ij}^k = relative utility for truck flows from state i to attraction state j ($j = 1-50$), and

$\hat{\alpha}_{0n}^k$, $n = 1, 2, \dots, 50$, $\hat{\alpha}_1^k$ and $\hat{\alpha}_2^k$ are OLS estimators of the coefficients α_{0n}^k , $n = 1, 2, \dots, 50$, α_1^k and α_2^k .

Calibration Results

The OLS regression was performed using the SPSS software to estimate the utility coefficients $\hat{\alpha}_{0n}^k$, $n = 1, 2, \dots, 50$, $\hat{\alpha}_1^k$, and $\hat{\alpha}_2^k$ for each commodity category k . To analyze the significance of the independent variables in predicting the dependent variable, statistical significance testing of the parameters was performed using the two-tailed Student's t test at a 90 percent confidence level. The final output of the OLS regression analysis for the utility functions of the production flow distribution model for truck flows from production state i to Texas by commodity is summarized in Table 4.1.

From the t statistic values in Table 4.1, it is evident that the variables (i.e., distance and relative attractions) are highly significant in the regression. Also, the fact that the adjusted R^2 values were found to be between 0.3 and 0.5 seems to suggest that the predictive power of the model can be considered reasonable.

Table 4.1 Outputs of the OLS Regression Analysis for the Production Flow Distribution Model

Commodity Category	Significant Variables	OLS Coefficient Estimates	Standard Error	<i>t</i> statistic	Adjusted R^2
Agriculture ($k = 1$)	Constant	1.504	0.560	2.6857	0.304
	d_{iT}	-0.003	0.000105	-28.479	
Food ($k = 2$)	d_{iT}	-0.004	9.8643E-05	-40.550	0.489
	FA_T^2	0.605	0.030	20.098	
Building materials ($k = 3$)	Constant	-1.597	0.510	-3.130	0.448
	d_{iT}	-0.003	7.516E-05	-39.913	
	FA_T^3	0.557	0.040	13.847	
Raw materials ($k = 4$)	Constant	1.117	0.448	2.496	0.418
	d_{iT}	-0.002	7.8E-05	-25.626	
	FA_T^4	0.175	0.030	5.796	
Chemicals and petroleum ($k = 5$)	d_{iT}	-0.003	7.85E-05	-38.191	0.423
	FA_T^5	0.444	0.026	16.995	
Wood ($k = 6$)	Constant	3.502	0.500	7.008	0.519
	d_{iT}	-0.004	9.23E-05	-43.307	
	FA_T^6	0.165	0.053	3.104	
Textiles ($k = 7$)	d_{iT}	-0.003	9.56E-05	-31.368	0.408
	FA_T^7	0.720	0.040	17.955	
Machinery ($k = 8$)	Constant	1.062	0.462	2.299	0.429
	d_{iT}	-0.003	9.798E-05	-30.617	
	FA_T^8	0.260	0.028	9.386	
Miscellaneous ($k = 9$)	d_{iT}	-0.003	0.0001	-29.987	0.295
	FA_T^9	0.128	0.025	5.113	

4.4.2 MNL Attraction Flow Distribution Model

In the MNL attraction flow model, the deterministic utility for truck flows of commodity k to attraction state j from production state i ($i = 1-50$; V_{ji}^k) is given by Equation 45:

$$V_{ji}^k = \beta_{0i}^k + \beta_1^k d_{ji} + \beta_2^k FP_i^k \quad (45)$$

where

β_{0i}^k , β_1^k and β_2^k are the coefficients to be determined during model calibration,
 d_{ji} is the centroidal distance between states j and i , and
 FP_i^k is the relative production level of states i for commodity group k .

As in the case of the production flow distribution model, the linear utility equation can be expressed as Equation 46:

$$V_{ji}^k = \beta_{01}^k X_{01} + \beta_{02}^k X_{02} + \dots + \beta_{050}^k X_{50} + \beta_1^k d_{ji} + \beta_2^k FP_i^k \quad (46)$$

where

X_{0m} , $m = 1, 2, \dots, 50$ denote dummy variables with values given by the following:

$X_{0m} = 1$ if $i = m$

$X_{0m} = 0$ if $i \neq m$

Because the dependent variables are in terms of the relative utilities computed from the CFS data, the coefficients of the linear utility functions for the attraction flow distribution model are estimated by performing linear regression of V_{ji}^k with X_{0m} , $m = 1, 2, \dots, 50$, $d_{ji|r}$ and $FP_{i|r}^k$, where $d_{ji|r}$ and $FP_{i|r}^k$ are as defined during model calibration phase. The OLS linear regression equation for the utility functions is given by Equation 47:

$$V_{ji}^k = \hat{\beta}_{01}^k X_{01} + \hat{\beta}_{02}^k X_{02} + \dots + \hat{\beta}_{050}^k X_{50} + \hat{\beta}_1^k d_{ji|r} + \hat{\beta}_2^k FP_{i|r}^k \quad (47)$$

where

V_{ji}^k = relative utility for truck flows to state j from production state i ($i = 1-50$), and

$\hat{\beta}_{0m}^k$, $m = 1, 2, \dots, 50$, $\hat{\beta}_1^k$ and $\hat{\beta}_2^k$ are OLS estimators of the coefficients β_{0m}^k , $m = 1, 2, \dots, 50$, β_1^k and β_2^k respectively.

The final output of the attraction flow distribution model calibration for the utility function for flows to attraction state j from Texas is summarized in Table 4.2. Similar to the results in Table 4.1, the t statistic values in Table 4.2, suggests that the variables (i.e., distance and relative productions) are highly significant in the regression. Also, with the exception of agriculture and miscellaneous commodities, most of the adjusted R^2 values were found to be between 0.4 and 0.5, which suggests that the predictive power of the model can be considered reasonable.

Table 4.2 Outputs of the OLS Regression Analysis for the Attraction Flow Distribution Model

Commodity Category	Significant Variables	OLS Coefficient Estimates	Standard Error	<i>t</i> statistic	Adjusted R^2
Agriculture ($k = 1$)	Constant	1.185	0.547	2.168	0.243
	d_{iT}	-0.002	8.9E-05	-22.467	
	FP_T^1	0.126	0.028	4.414	
Food ($k = 2$)	d_{iT}	-0.003	8.218E-05	-36.502	0.435
	FP_T^2	0.504	0.031	16.520	
Building materials ($k = 3$)	d_{iT}	-0.004	8.986E-05	-44.510	0.521
	FP_T^3	0.409	0.031	13.319	
Raw materials ($k = 4$)	d_{iT}	-0.002	8.857E-05	-22.580	0.191
	FP_T^4	0.104	0.028	3.649	
Chemicals and petroleum ($k = 5$)	d_{iT}	-0.003	8.603E-05	-34.871	0.390
	FP_T^5	0.371	0.028	13.205	
Wood ($k = 6$)	Constant	1.060	0.457	2.319	0.508
	d_{iT}	-0.004	9.249E-05	-43.244	
	FP_T^6	0.493	0.035	14.050	
Textiles ($k = 7$)	d_{iT}	-0.003	9.203E-05	-32.598	0.444
	FP_T^7	0.431	0.033	13.167	
Machinery ($k = 8$)	Constant	1.525	0.503	3.034	0.504
	d_{iT}	-0.003	8.185E-05	-36.650	
	FP_T^8	0.096	0.041	2.360	
Miscellaneous ($k = 9$)	d_{iT}	-0.003	0.000106	-28.300	0.276
	FP_T^9	0.168	0.025	6.685	

4.5 Calculation Of County-Level Truck Flows

As indicated earlier, the calibrated state-level MNL models for the production and attraction flow distributions can be used to generate Texas county-to-state, state-to-Texas county, and Texas county-to-county truck flows from the available CFS data.

4.5.1 Texas County-to-State flows

To illustrate the calculations involved in determining county-to-state flows by commodity group, the disaggregation procedure to determine the fraction of total truck food tonnage to

Colorado from Texas, originating from Harris County (Texas), is discussed below. Truck flows of food commodities from Harris County to Colorado can be graphically represented as follows in Figure 4.1:

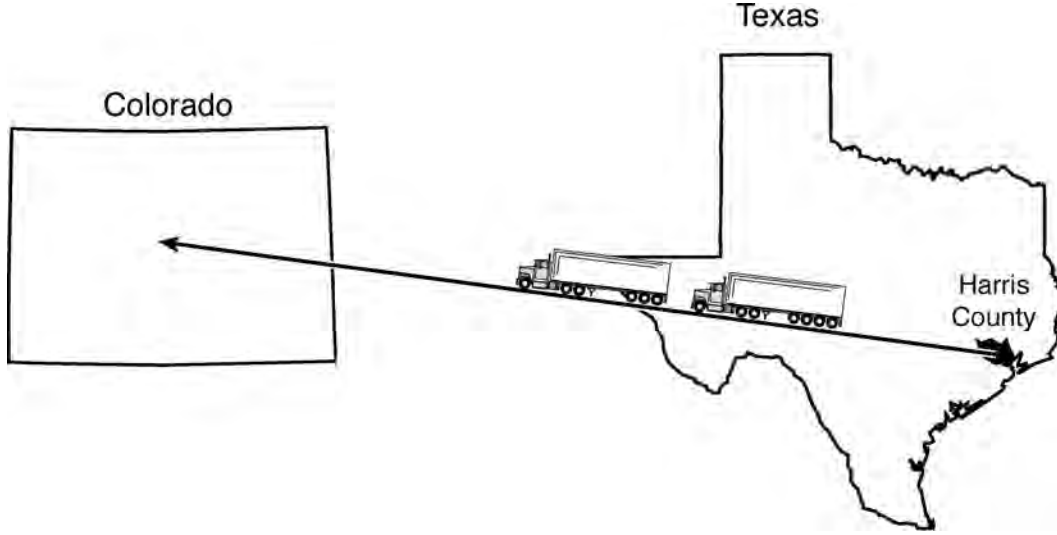


Figure 4.1 Truck flows of food commodities from Harris County to Colorado

According to the CFS, 585,000 tons of food moved by truck to Colorado from Texas.

$$TA_{CO, TX}^{Food} = 585,000 \text{ tons}$$

Using the attraction flow distribution model, the fraction of the total food tonnage originating in Harris County can be calculated as follows:

$$T_{CO, i=HarrisCounty}^{Food} = TA_{CO, TX}^{Food} * \frac{e^{V_{CO, i=HarrisCounty}^{Food}}}{\sum_{i=1}^{254} e^{V_{CO, i}^{Food}}} \quad (48)$$

The calibrated utility function for truck flows of food to Colorado from Harris County is given by Equation 49:

$$V_{CO, i=HarrisCounty}^{Food} = (-0.003 * d_{CO, HarrisCounty}) + (0.504 * FP_{HarrisCounty}^{Food}) \quad (49)$$

where

$$d_{CO, HarrisCounty} = 1000.06 \text{ miles, and}$$

$$FP_{HarrisCounty}^{Food} = \frac{P_{HarrisCounty}^{Food}}{\sum_{i=1}^{254} P_i^{Food}} * 100 = 16.37\%$$

Substituting in Equation 49 results in Equation 50.

$$V_{CO,i=HarrisCounty}^{Food} = (-0.003 * 1000.06) + (0.504 * 16.367) = 5.25 \quad (50)$$

Summation of the exponential of the utilities across all production counties gives Equation

51.

$$\sum_{i=1}^{254} e^{V_{CO,i}^{Food}} = 338.77 \quad (51)$$

Entering the values into Equation 48 provides the total annual truck food tonnage to Colorado from Harris County:

$$T_{CO,i=HarrisCounty}^{Food} = 585,000 * \frac{190.3665}{338.7739} = 328,728 \text{ tons.} \quad (52)$$

4.5.2 State-to-Texas County Flows

To illustrate the calculations involved in determining state-to-Texas county truck flows by commodity group, the disaggregation procedure to determine the fraction of total truck food tonnage from Colorado to Texas, destined to Harris County, is discussed below. Annual truck flows of food commodities from Colorado to Harris County, Texas, can be graphically represented as follows in Figure 4.2:

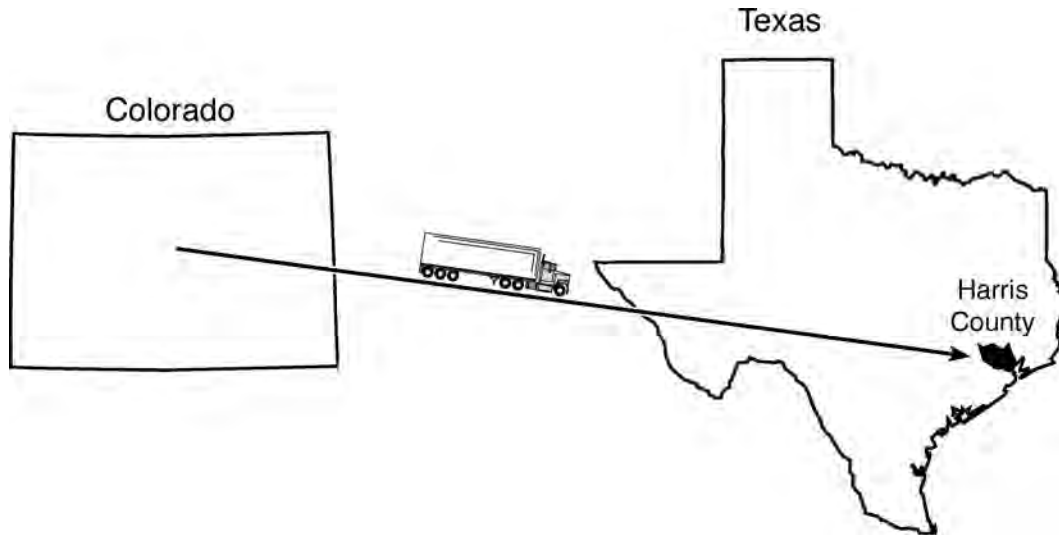


Figure 4.2 Truck flows of food commodities from Colorado to Harris County

According to the CFS, 299,000 tons of food moved by truck from Colorado to Texas in 1997:

$$TP_{CO,TX}^{Food} = 299,000 \text{ Tons}$$

Using the production flow distribution model, the fraction of the total food tonnage by truck from Colorado to Texas destined for Harris County can be determined as follows:

$$T_{CO,j=HarrisCounty}^{Food} = TP_{CO,TX}^{Food} * \frac{e^{V_{CO,j=HarrisCounty}^{Food}}}{\sum_{j=1}^{254} e^{V_{CO,j}^{Food}}} \quad (53)$$

The calibrated utility function for truck flows of food commodities from Colorado to Harris County is given by Equation 54:

$$V_{CO,j=HarrisCounty}^{Food} = (-0.004 * d_{CO,HarrisCounty}) + (0.605 * FA_{HarrisCounty}^{Food}) \quad (54)$$

where

$$d_{CO,HarrisCounty} = 1000.06 \text{ miles}$$

$$FA_{HarrisCounty}^{Food} = \frac{A_{HarrisCounty}^{Food}}{\sum_{j=1}^{254} A_j^{Food}} * 100 = 18.65\%$$

Substituting these values in Equation 54 results in Equation 55:

$$V_{CO,j=HarrisCounty}^{Food} = (-0.004 * 1000.06) + (0.605 * 18.65064) = 7.28. \quad (55)$$

Summation of the exponential of the utilities across all attraction counties gives Equation 56:

$$\sum_{j=1}^{254} e^{V_{CO,j}^{Food}} = 1560.67. \quad (56)$$

Entering the values into Equation 53 provides the total annual truck food tonnage moved from Colorado to Harris County, Texas:

$$T_{CO,j=HarrisCounty}^{Food} = 299,000 * \frac{1455.923}{1560.665} = 278,933 \text{ tons}. \quad (57)$$

4.5.3 Texas County-to-County Truck Flows

Texas county-to-county truck flows were calculated for nine commodity categories from the Texas intrastate truck flow data captured in the CFS. To illustrate the intercounty truck flow disaggregation procedure, the computation of annual truck flows of food from Harris County to Dallas County is presented below and can be graphically illustrated as follows in Figure 4.3.

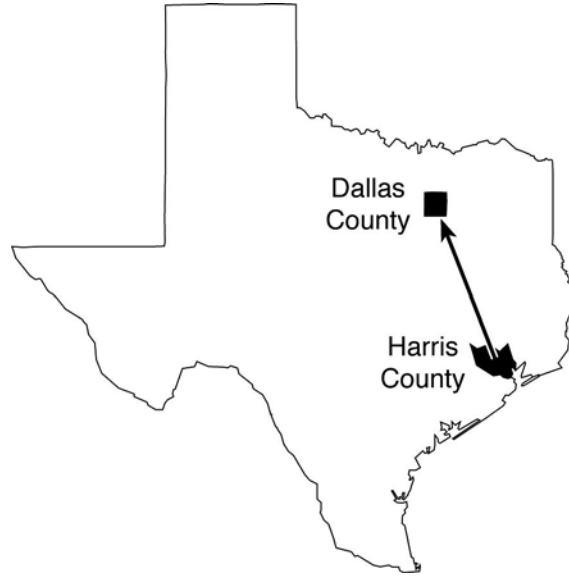


Figure 4.3 Truck flows of food commodities from Harris County to Dallas County

According to the CFS, 28,288,000 tons of food commodities moved intrastate in Texas in 1997:

$$T_{TX,TX}^{Food} = 28,288,000 \text{ tons.}$$

Using the attraction flow distribution model, the fraction of the total intrastate truck food flows originating in Harris County can be determined by Equation 58:

$$T_{i=HarrisCounty,TX}^{Food} = T_{TX,TX}^{Food} * \frac{e^{V_{i=HarrisCounty,TX}^{Food}}}{\sum_{i=1}^{254} e^{V_{i,TX}^{Food}}} \quad (58)$$

The calibrated utility function for food flows from Harris County to Texas is shown in Equation 59:

$$V_{i=HarrisCounty,TX}^{Food} = (-0.003 * d_{HarrisCounty,TX}) + (0.504 * FP_{HarrisCounty}^{Food}) \quad (59)$$

where

$$d_{HarrisCounty,TX} = 309.29 \text{ miles.}$$

$$FP_{HarrisCounty}^{Food} = \frac{P_{HarrisCounty}^{Food}}{\sum_{i=1}^{254} P_i^{Food}} * 100 = 16.37\%.$$

Substituting these values in Equation 59 results in Equation 60:

$$V_{i=HarrisCounty,TX}^{Food} = (-0.003 * 309.29) + (0.504 * 16.367) = 7.32 . \quad (60)$$

Summation of the exponential of the utilities across all production counties gives Equation 61:

$$\sum_{i=1}^{254} e^{V_{i,TX}^{Food}} = 2317.08 \quad (61)$$

Entering the values into Equation 58 provides the total annual truck food tonnage from Harris County to Texas:

$$T_{i=HarrisCounty,TX}^{Food} = 28,288,000 * \frac{1511.863}{2317.08} = 18,457,533 \text{ tons.} \quad (62)$$

Using the production flow distribution model, the fraction of the intrastate truck food tonnage originating in Harris County destined for Dallas County can be determined as follows:

$$T_{i=HarrisCounty,j=DallasCounty}^{Food} = T_{i=HarrisCounty,TX}^{Food} * \frac{e^{V_{i=HarrisCounty,j=DallasCounty}^{Food}}}{\sum_{j=1}^{254} e^{V_{i=HarrisCounty,j}^{Food}}} \quad (63)$$

where

$$V_{i=HarrisCounty,j=DallasCounty}^{Food} = (-0.004 * d_{HarrisCounty,DallasCounty}) + (0.605 * FA_{DallasCounty}^{Food}), \quad (64)$$

$$d_{HarrisCounty,DallasCounty} = 229.77 \text{ miles, and}$$

$$FA_{DallasCounty}^{Food} = \frac{A_{DallasCounty}^{Food}}{\sum_{j=1}^{254} A_j^{Food}} * 100 = 12.42\%.$$

Substituting these values in Equation 64 results in Equation 65:

$$V_{i=HarrisCounty,j=DallasCounty}^{Food} = (-0.004 * 229.77) + (0.605 * 12.41753) = 6.59. \quad (65)$$

Summation of the exponential of the utilities across all attraction counties gives Equation 66:

$$\sum_{j=1}^{254} e^{V_{i=HarrisCounty,j}^{Food}} = 80397.76. \quad (66)$$

Entering the values into Equation 63 provides the total annual food truck tonnage moving from Harris County to Dallas County:

$$T_{i=HarrisCounty,j=DallasCounty}^{Food} = 18,457,533 * \frac{730.351}{80397.7607} = 167,672 \text{ tons.}$$

4.6 Strengths and Limitations of the Modeling Methodology

4.6.1 Strengths of the Modeling Methodology

The MNL modeling approach, developed in this research, provides a cost-effective methodology for generating disaggregate truck flow data from the publicly available CFS data. The following are considered the strengths of this approach:

- Ease of model calibration. MNL production and attraction flow distribution models of truck movements can be relatively easily calibrated using the publicly available truck data from the CFS. The utilities for truck flows can be modeled as linear functions of the explanatory variables that impact truck flow distribution, and the coefficients can be estimated using simple OLS regression analysis.
- Predictive power of model. The degree to which the model predicts the observed truck distribution attests to the accuracy and reliability of the model output. Adjusted R^2 values measure the predictive power of the model. The adjusted R^2 values were found to be between 0.4 and 0.5 in the majority of the cases (see Tables 4.2 and 4.1), which are reasonable for freight models considering the limitations of the data used.
- Cost effective. The models were calibrated using publicly available truck flow data captured in the CFS. This approach is thus very cost effective.
- Information on data accuracy and reliability. Because the inherent assumptions, limitations, and margins of error associated with the CFS data are well documented, the accuracy and reliability of the disaggregated data generated from the model are known.

4.6.2 Limitations of the Modeling Methodology

A MNL modeling approach was used to disaggregate the state data to the county level. Although MNL models are particularly suited for modeling the production and attraction flow distribution of trucks and can be calibrated relatively easily, their application is impacted by the quality and reliability of the data used and the power of the explanatory variables in predicting the distribution of truck flows. This section highlights the specific limitations of the methodology.

Limitations of the CFS

The CFS does not report interstate truck flow data in the following situations.

- Magnitude of truck flows. Truck flows less than one unit of measure are not reported in the CFS database.

- Sampling variability in data collection. The CFS database is developed by conducting surveys of a sample of establishments in the mining, manufacturing, wholesale, and certain retail industries. The data are subject to sampling variability because only a subset of the entire population is surveyed. To ensure the reliability of the reported data, estimates with sampling variability higher than a certain value are not reported in the CFS.
- Federal disclosure rules for Census Bureau reports. Federal law data states that any data that would disclose the operations of a specific firm or establishment cannot be reported in the CFS.
- Industry coverage. The CFS covers only shipments originating from the manufacturing, mining, wholesale, and selected retail sectors. The farming, forestry, fishing, construction, crude petroleum production, household, government, foreign establishment, and most retail and service businesses are thus not covered (Transportation Research Board, 2003). Inadequate interstate truck flow data for primary agriculture movements, secondary traffic, and hazardous material (HAZMAT) shipments thus have prevented the development of truck flow distribution models for these shipments.
- Data coverage. Because the CFS does not provide routing details for shipments between origins and destinations, external–external (through) truck flows in Texas could not be estimated (Cambridge Systematics, Inc. et. al. 1997).

Explanatory Variables Included in the Modeling Approach

The MNL model for production (attraction) flows is calibrated with the assumption that the distribution of truck flows is a function of the interstate centroidal distances and attraction levels (production levels) of states. These assumptions are prone to the following limitations.

- Contribution of other factors. A number of factors other than distance and the production/attraction levels of states can impact the truck flow distribution between states, including (a) long-standing trading relationships between states, (b) the location of intermodal terminals, inland ports, and seaports, and (c) a number of mode choice factors. Because these factors (explanatory variables) have not been considered in the model, it does affect the accuracy of the results.
- Centroidal distances. The impedance measure for truck flows included in the model is represented by the interstate centroidal distances. Truck flows, however, typically occur between regions of economic activity, which are not necessarily located at the geometrical centroid of the state. Centroidal distance is thus a crude proxy for the impedance measure for interstate truck flows.

4.7 Comparison of Model Output with Reebe TRANSEARCH Database

One of the primary data sources in compiling the Reebe TRANSEARCH database is the CFS data. The authors thus compared the model estimates with the TRANSEARCH data for Texas to determine whether there is any statistically significant difference between the model estimates and the Reebe data used in the SAM. Because the model estimates and the TRANSEARCH data are not from completely independent sources, the paired sample t test was employed to compare the means of the two data sets.

The paired sample t test is used to compare the means of two dependent sets of samples. In comparing the means of the two data sets, the individual fields representing truck flows (in tons) between origin–destination pairs are thus not randomly selected from the population. In other words, comparisons are made for origin–destination truck flows that are reported in both the databases. The paired sample t test thus determines the differences in the paired data and reports the probability that the mean of the paired differences is zero, given a specified confidence interval. The null hypothesis being tested is that the mean of the paired differences is equal to zero:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

where

H_0 = null hypothesis,

H_1 = alternate hypothesis, and

μ_d = mean of paired differences.

The SPSS statistical software was used to conduct the paired sample t test. The software calculates the mean of the paired differences, the standard deviation of the paired differences, and the t statistic, which is equal to the mean divided by the standard error of the mean. A confidence interval of 95 percent was specified. For this confidence interval, the critical t value is equal to 1.645. Thus, if the absolute value of the t statistic computed from the paired sample t test is greater than the critical t value of 1.96, the null hypothesis is rejected at the 95 percent confidence level (two-tailed test). This implies that the mean of the paired differences is significantly different from zero at the 95 percent confidence level. Alternatively, if the absolute value of the t statistic is less than 1.96, the null hypothesis is accepted. This implies that the mean of the paired differences is statistically equal to zero at the 95 percent confidence level. The following tables summarize the statistical results of the paired sample t tests that were conducted.

From Table 4.3 it is evident that for six out of the nine commodity groups, the mean of the model estimates for Texas county-to-state truck flows is statistically equal to the mean of the TRANSEARCH data for these flows, which implies that the model results are significantly similar to the TRANSEARCH data. A statistically significant difference between the model estimates and the TRANSEARCH data was observed for the other commodity groups. This can partly be explained by the fact that the TRANSEARCH database does not capture truck flows for many Texas counties in terms of these commodity groups.

Table 4.3 Paired Sample t Test Results for the Texas County-to-State Truck Flow Comparison

Commodity	Paired Sample t Statistic	Critical t (95% CI)	Null Hypothesis ($\mu_d = 0$)
Agriculture	16.329	1.96	Reject
Food	1.837	1.96	Accept
Building materials	1.399	1.96	Accept
Raw materials	4.271	1.96	Reject
Chemicals and petroleum	1.052	1.96	Accept
Wood	-0.727	1.96	Accept
Textiles	1.722	1.96	Accept
Machinery	1.881	1.96	Accept
Miscellaneous	5.616	1.96	Reject

Note: $n_1 = n_2 = 12,700$

Table 4.4 shows that in the case of the state-to-Texas county truck flows, the model estimates are statistically similar to the TRANSEARCH data for four of the nine major commodity groups.

Table 4.4 Paired Sample t Test Results for the State-to-Texas County Truck Flow Comparison

Commodity	Paired Sample t Statistic	Critical t (95% CI)	Null Hypothesis ($\mu_d = 0$)
Agriculture	1.371	1.96	Accept
Food	4.708	1.96	Reject
Building materials	-4.845	1.96	Reject
Raw materials	0.936	1.96	Accept
Chemicals and petroleum	-0.229	1.96	Accept
Wood	-1.661	1.96	Accept
Textiles	4.592	1.96	Reject
Machinery	-2.794	1.96	Reject
Miscellaneous	6.346	1.96	Reject

Note: $n_1 = n_2 = 12,700$

For Texas county-to-county truck flows, the model estimates are statistically similar to the TRANSEARCH data for five of the nine major commodity groups (see Table 4.5).

Table 4.5 Paired Sample t Test Results for the Texas County-to-County Truck Flow Comparison

Commodity	Paired Sample t Statistic	Critical t (95% CI)	Null Hypothesis ($\mu_d = 0$)
Agriculture	325.388	1.96	Reject
Food	0.398	1.96	Accept
Building materials	0.155	1.96	Accept
Raw materials	6.735	1.96	Reject
Chemicals and petroleum	0.483	1.96	Accept
Wood	-4.604	1.96	Reject
Textiles	-1.323	1.96	Accept
Machinery	0.483	1.96	Accept
Miscellaneous	65.715	1.96	Reject

Note: $n_1 = n_2 = 64,516$

4.8 Concluding Remarks

This section of the report presented a novel approach for estimating Texas county-level truck commodity flow data from the state-to-state freight flow data captured by the publicly available CFS database. It is believed that the MNL approach, despite the limitations highlighted, provides a cost-effective means of estimating county-level truck flows for Texas from readily available public freight data sources for at least some of the commodity groups. A comparison with the Reebie data revealed that for approximately half of the commodity groups, the model estimates are statistically similar to the TRANSEARCH data. Therefore, although the disaggregate Texas county flows generated from the model suffer from the inherent limitations of the CFS (sampling error, industry coverage, and missing data), the model can be used in the short term to generate county-level data.

5. Truck Travel Survey Methods

This section of the report provides an overview of the available primary freight data collection methods that have been used and highlights two data collection approaches that showed the most promise of providing the Texas Department of Transportation (TxDOT) with statistically reliable and verifiable information on truck commodity flows required for the Statewide Analysis Model (SAM). For additional detailed information about these approaches, the reader is referred to a document titled *Texas Truck Data Collection Guidebook*, which was prepared as part of this TxDOT research project.

5.1 Truck Travel Survey Methods

Primary freight data collection efforts involve collecting freight flow data directly from the freight community (i.e., shippers, carriers, receivers, and freight forwarders) through surveys, including roadside intercept surveys, mail-out/mail-back questionnaires, combined telephone–mail-out/mail-back questionnaires, and telephone interviews. Done correctly, these survey methods are, in general, the most reliable and accurate methods of obtaining freight flow data for statewide freight planning programs. Collecting primary freight data from surveys of shippers and carriers is, however, a costly and time-consuming process, especially when conducted at the state level. These survey data collection methods and their most significant strengths and weaknesses are highlighted in Table 5.1.

Table 5.1 Summary of Truck Travel Survey Methods

Survey Methods	Typical Completed Surveys (% of Total Population)	Typical Response Rate	Strengths	Weaknesses
Telephone Interviews	4–5%	40–50%	<p>Easy to implement</p> <p>High response rate</p> <p>Easy to follow up</p> <p>No disruption of traffic</p> <p>Quicker turnaround than mail</p> <p>Generally good information from those that respond</p>	<p>Difficulty obtaining appropriate and correct telephone numbers</p> <p>Can call only during business hours</p> <p>“Phone tagging” problem</p> <p>Limited time on phone if respondent is busy</p> <p>The number and skill of the interviewers may make this method costly</p> <p>Requires access to vehicle registration file</p> <p>Under-representation of out-of-state trucks in sampling frame</p>
Mail-out/Mail-back	1–5%	10–45%(1)	<p>Less costly</p> <p>Easy to implement</p> <p>No disruption of traffic</p> <p>Good response rate with certified mail, but in general lower response rate</p> <p>Generally good information from those that respond</p> <p>Only follow-up of nonresponses necessary</p>	<p>Low overall response rate</p> <p>Low response from small truck owners</p> <p>Under-representation of out-of-state trucks in sampling frame; thus, low response from out-of-state trucks</p> <p>Need to follow-up on nonresponses</p> <p>Reliability and completeness depend on finding the appropriate individual within an organization or company</p> <p>Possible bias owing to better response from drivers/owners</p> <p>Requires access to registration file</p>
Combined Telephone–Mail-out/Mail-back	3–10%	20–80%(2)	<p>Improved response rate over mail-out/mail-back alone</p> <p>Easy to implement</p> <p>No disruption of traffic</p> <p>Early identification of owners who agree to participate and [decrease in] potential nonresponses through phone contact</p> <p>Generally good information from those that respond</p> <p>Improved ability to explain questions and</p>	<p>Difficulty obtaining appropriate and correct telephone numbers</p> <p>Can only call during business hours</p> <p>“Phone tagging” problem</p> <p>Limited time on phone if respondent is busy</p> <p>Relatively low response</p> <p>High cost of telephone follow-ups</p> <p>Requires access to registration file</p> <p>Under-representation of out-of-state trucks in sampling frame</p>

Survey Methods	Typical Completed Surveys (% of Total Population)	Typical Response Rate	Strengths	Weaknesses
Roadside Intercept/Interview	8–35% (3)	95–100%	<p>clarify intent, leading to better data detail</p> <p>Phone contact may help adjust sample size for mail-out/mail-back</p> <p>Complete information (i.e., origin–destination, route, trip purpose, specific commodity, etc.)</p> <p>High response rate</p> <p>Better sampling control</p> <p>Good representative sample of trucks entering or leaving a cordon line (i.e., including vehicles passing through from outside geographical area)</p> <p>Easy comparison with mainstream traffic through field counts at survey location</p> <p>Ability to expand to total truck traffic population</p>	<p>High labor requirement</p> <p>Potentially costly</p> <p>Potential disruption to traffic flows</p> <p>Quality and conduct of survey affected by weather, lighting, time of day</p> <p>Hazardous to survey crew</p> <p>Time constraint</p> <p>No follow-up possible</p> <p>Enforcement problems</p> <p>Drivers avoiding the survey station</p> <p>Limited locations where survey may be implemented</p> <p>Only represents trucks traveling on road along survey station, not entire region</p>
Personal Interviews			Complete information	<p>Most costly method of conducting surveys</p> <p>Generally involves a smaller, more select (or targeted) sample</p> <p>High labor requirement</p> <p>Not appropriate for population of statewide truck travel databases</p>

(1) The higher response rate was due to better survey participation from large truck fleet operators.

(2) The higher response rate was due to an employer survey conducted in California (1991 Caltrans–Alameda County Survey)

(3) The higher percentage is from the 1988 Ontario survey that surveyed fifty-seven locations over a 1,855-hour period throughout Ontario Province.

Source: Adapted from Lau, 1994; NCHRP Report 388, 1997; Federal Highway Administration, 1996; and Jessup, Casavant, and Lawson, 2004.

5.2 Recommended Truck Data Collection Approaches

This section of the document highlights two data collection approaches (i.e., roadside intercept surveys and truck carrier participation programs) that showed the most promise of providing TxDOT with statistically reliable and verifiable information on the truck commodity flows required for the SAM.

5.2.1 Roadside Intercept Surveys

Despite the weaknesses associated with roadside intercept surveys highlighted in Table 5.1, this method of truck data collection has proven to yield statistically reliable data on statewide intercity truck flows. TxDOT has considerable experience in conducting external station passenger and commercial vehicle surveys. Detailed specifications exist for conducting these surveys (see text box). The objective of the guidance provided in this section of the document is thus not to replace, contradict, or replicate the specifications used by TxDOT. Rather, it is anticipated that the information provided can be used as a framework to modify the specifications in the future when a decision is made to collect primary data to populate the SAM database.

5.2.2 Data Collection and Sampling

Because it is impossible to obtain data from every single truck that travels in or through Texas, inferences or predictions are typically made from an appropriately selected sample. Sampling typically results in savings in resources, money, and time. Incorrect sampling procedures or human judgment may, however, cause bias and result in the collection of unreliable data (Snedecor, 1989). The objective of this section is to provide guidance on sampling to answer three questions:

- Where to sample? In other words, how to select appropriate sites where roadside intercept surveys can be conducted?
- Who to sample? In other words, which trucks at the sites?
- When to sample? In other words, which days of the week and seasons to account for seasonal variation?

TxDOT Requirements for Conducting External Surveys

The following components are specifically described in TxDOT's specifications for conducting external passenger and commercial vehicle surveys (TxDOT, 2003):

- Demonstration external station travel survey. This component has the objective to inform the news media, local officials, and the public of how the surveys will be conducted.
- External station travel survey (pilot). This component has the objective to test questionnaires, evaluate surveyor training, traffic control plans, and the survey approach.
- External station travel survey (regular). TxDOT specifies the external stations for the area, the interview method, the direction of survey, the questionnaire, the times of survey, the management of survey queues, and the method, day, and extent of the vehicle classification counts. In addition, TxDOT is very specific about how survey data should be coded and entered, how errors or problems with the survey data should be corrected and the consequences of failure to do so, how data should be processed and geocoded, and finally how the survey should be documented and reported.
- Vendor personnel. TxDOT specifies the requirements in terms of personnel conduct, sexual harassment training, and the type of equipment that is allowed on site.
- Trained interviewers and supervisors. TxDOT specifies the minimum training for interviewers and supervisors, the dress code for interviewers, that the transportation needs of interviewers and field staff be met, and that breaks and rest periods are scheduled for all interviewing staff.
- Traffic control. TxDOT specifies that an experienced (of at least three years) firm or corporation should provide traffic control services. The vendor must conduct traffic control in accordance with applicable federal, state, and local government regulations. Requirements for (a) obtaining a written release from the TxDOT district representative, (b) equipment based on the roadway facility, (c) changeable message boards, (d) language proficiency of field coordinator, (e) personnel and equipment necessary for installation and removal of traffic control equipment, (f) law enforcement, and (g) a videotape of the traffic control set-up before each survey are specified.
- Flagging. TxDOT specifies that all flaggers should be trained by someone who has completed flagger-training courses by, for example, the Texas Engineering Extension Service (TEEX) or the American Traffic Safety Services Association (ATSSA). In addition, TxDOT specifies that the vendor provides trained flaggers at each site, that the equipment used by flaggers (i.e., safety hard hats, safety vests, and steel-toed footwear) be approved by the American National Standards Institute (ANSI) and meet the specifications for quality and visibility used by TxDOT as applicable, and finally the dress code for flaggers and traffic control personnel.
- Bilingual requirements. To accommodate Hispanic respondents, TxDOT requires that Spanish versions of the survey forms be developed and implemented and that bilingual interviewers and flagging/traffic control staff be used.
- Status reports/deliverables. Monthly status reports are required. All deliverables and products are clearly specified.
- Finally, TxDOT also specifies how travel and per diem should be accounted for, how complaints filed by the public against the vendor should be handled, how liquidated damages will be assessed, when the purchase order will be canceled, the requirements for subcontractors, record keeping, federal requirements and other specifications pertaining to the participation of disadvantaged business enterprises, historically underutilized businesses, Title VI requirements, and Executive Order 11246 entitled *Equal Employment Opportunity*, and compliance with laws.

Site Selection: Identification of Geographically Dispersed Sites

The optimal survey sample is influenced not only by the level of reliability that is required, but also the costs of implementing the survey, the number of interviewers required, the time frame, the safety of the interviewers, and the impacts to the general public. In identifying the physical roadway sections for the roadside intercept surveys, the following issues should be taken into consideration:

- Texas's eight National Transportation Analysis Regions (NTARs)⁷, as defined by the US Department of Transportation, to ensure geographical coverage of the entire state;
- non-urban roadway sections to capture county-to-county movements;
- roadway functional system to capture truck movements on all major road classes; and
- traffic volume and vehicle classification to ensure that a significant volume of daily truck traffic traverses the roadway section.

To ensure geographical coverage of the state of Texas and to account for variations in the characteristics of trucks that use different roadways, it is recommended that ninety-six sites at a minimum should be identified: three sites per major highway type (i.e., interstate, state highway, Texas highway, and farm-to-market/ranch-to-market) in each of the eight NTARs of Texas. In selecting the location of the survey stations, care should be taken to ensure that a significant proportion of the interurban freight trips in the NTAR can be captured in an average week.

The vendors should visit each of the roadway sections to determine an appropriate site where surveys can be conducted (e.g., availability of weigh stations, truck rest stops, etc.). Attention should be paid to existing roadway conditions, sight distances, prevailing vehicle speeds, and the presence of shoulders and auxiliary lanes. After determining the feasibility of the sites as interview sites, the vendors are advised to share the survey plan with the TxDOT district offices having jurisdiction over the roadways at each site. Finally, Traffic Control Plans should be prepared, detailing the existing roadway geometry, the types of traffic control devices that will be used, and areas of refuge for the interviewers at roadside locations. These Traffic Control Plans need to be approved by TxDOT.

Who to Sample?

Although it seems reasonable to follow a systematic sampling approach where every k th vehicle is sampled, the objective should be to interview as many trucks as possible given available space and an interviewer to administer the survey. Minimum targets can be set—for example, to have an objective of surveying 10 percent of all trucks on interstate facilities, 5

⁷ For additional information as to how NTARs are defined, the reader is referred to the following internet site: http://www.bts.gov/programs/commodity_flow_survey/methods_and_limitations/national_transportation_analysis_regions/ntar_defined.html

percent of all trucks on major corridors (i.e., U.S. highways and state highways), and 50 percent of all trucks at sites with low truck traffic volumes.

When to Sample?

To account for seasonal variation, day-of-the week variation, and time-of-day variation, it is recommended that roadside intercept surveys be conducted at a minimum every season (four times a year) over a one-week period for 24 hours per day at each site. Surveying trucks for a 24 hour period is important to accurately capture time-of-day variation in truck movements. Given the number of sites proposed for Texas (i.e., ninety-six sites), TxDOT is advised to schedule the seasonal surveys over a two-month period. Depending on the location of the sites (e.g., the distance between sites), considerable care should be taken in scheduling the surveys to avoid surveying the same flow of trucks at more than one site.

If it is considered too expensive to conduct roadside intercept surveys seven days a week for 24 hours each day, an alternative is to collect truck data on fewer days. Wednesdays are typically chosen to collect truck travel data, because truck flows tend to be higher/lower toward the beginning and end of the week. Most holidays or observances of holidays also typically occur on Mondays or Fridays. It is, however, strongly recommended that intercept surveys be conducted at a minimum on one weekday and one weekend day to account for day of the week variations, because truck flows are typically lower on the weekends than on weekdays (Personal Communication with Mark Hodges, 2005).

Systematic Sampling

In systematic sampling, a starting point is arbitrarily selected. Then every k th element in the population is chosen to be in the sample (Triola, 2004). Although not considered a strictly random selection, the results will not necessarily be biased.

5.2.3 Recruiting Survey Personnel

The Strategic Freight Transportation Analysis (SFTA) study found that approximately fifteen to eighteen people were required at each site over a 24-hour period for conducting the interviews. This translates into an average of four to five people per interview shift per site (Clark, Jessup, and Casavant, 2002). The finding and hiring of a large number of short-term employees can translate into a major cost and thus concern in performing roadside intercept interviews (Gillis, 1994). In the past, TxDOT has entered into a contract with vendors, such as Graham Traffic or Alliance Texas, to conduct external station surveys for both passengers and commercial vehicles. These vendors have access to trained surveyors and supervisors, although the extent of the proposed effort might require the recruiting and training of additional surveyors and supervisors.

To reduce costs, it is possible to recruit volunteers from local organizations (e.g., Lions Club, Student Government) or educational institutions (e.g., local high schools or universities) who are interested in performing community service. Recruiting volunteers who reside close to the interview sites will reduce transportation costs, and their familiarity with the area will result in better knowledge and an understanding of the regions where the interviews will be conducted (Gillis, 1994). Assuming there are no groups that are willing to provide volunteer labor, an alternative is to allow groups to use the roadside intercept surveys as a fund-raising opportunity. This will cost significantly less than hiring a professional surveying company to perform the interviews. Although this might not be feasible in all situations, it is worth further consideration

in light of the fact that the Lions Club conducted roadside intercept surveys to raise funds for the Washington State freight truck origin and destination study to raise funds (Gillis, 1994).

5.2.4 Training Survey Personnel

Both classroom and on-site instruction of survey personnel is crucial to ensure that the data collected are ultimately reliable. The following points should be reviewed at the training sessions (Clark, Jessup, and Casavant, 2002, and Gillis, 1994).

- Project goals and objectives to enable surveyors to answer questions from truck drivers relating to the purpose of the data collection effort
- Detailed review of the interview questionnaire
- Identification and familiarization with truck and trailer configurations
- Personal interview techniques, including greeting, etiquette and proper behavior, and asking and phrasing of survey questions
- Personal safety, including letters of release of liability
- Personal conduct, including on-time performance, acceptable attire, writing legibly, and items to bring on-site

5.2.5 Survey Equipment Needs

Equipment needs will vary depending on the volume of truck traffic at the site and when the surveys are conducted. Besides general office supplies, other equipment, such as cones and safety vests, are required to ensure the safety of the survey crew. In addition, it is recommended that the surveyors be provided with attire (i.e., hard hats) that will enable truck drivers to identify the survey crew. The following materials checklist was used in the SFTA study (Clark, Jessup, and Casavant, 2002).

- Surveys, enough for each site
- Pens, highlighters, pencils, staplers, and staples
- Clipboards
- Reflective orange safety vests
- Traffic cones
- “Survey Team Ahead” sign
- Hard hats
- Plastic containers for storing surveys
- Headlamps for nighttime survey shifts

- Two-way radios
- Tally counters

5.2.6 Development of Relationship with Commercial Vehicle Enforcement Officers

The presence of uniformed officers at the interview site may prove to be beneficial. Cooperation from uniformed vehicle enforcement officers, employees of the Department of Public Safety, or police officers will, most importantly, ensure the safety of the survey crew. Secondly, many truck drivers may be intimidated by the presence of a uniformed officer in the area. They may be concerned that they were being pulled over for a violation. After learning about the survey, these drivers might be relieved and more willing to participate. This will enhance participation and result in more data (Gillis, 1994).

5.2.7 Public Notification

The public needs to be notified about the goals and purpose of the data collection effort (Jessup, Casavant, and Lawson, 2004). Public awareness campaigns should be designed and implemented to inform the trucking community about the purpose of the survey and the potential benefits of participating well in advance of the actual survey dates.

5.2.8 Survey Design

Survey instruments must be developed in a manner that ensures that participants are not burdened or overwhelmed by the interview process. Also, questions must be worded appropriately to ensure correct responses and improve response rates. For the SAM truck travel database, data are required regarding the vehicle configuration, commodity origin and destination, cargo/commodity type, cargo weight, commodity value, and operational characteristics (i.e., truckload, less-than-truckload, or empty).

When developing survey questions, it is important to keep in mind that truck driver cooperation is crucial to the success of the survey. To minimize the inconvenience to the truck driver, it is important to keep the survey instrument as brief as possible. This also allows more trucks to be surveyed. Some questions could thus be completed by the surveyor through observation. Examples of such questions include the date, time of day, and the vehicle configuration (see survey instrument). This will reduce the amount of time spent conducting the actual interview with the participating truck driver.

All questions should avoid wording bias. Examples of wording bias would be “loaded” questions or structuring questions in a way to elicit a certain response. Answers to questions should be kept as simple as possible by minimizing the amount of writing necessary to complete the survey. Checkboxes for frequently appearing answer choices can be used to save time.

In addition, it is important that the interviewer phrase the questions properly to the participating truck driver to avoid any confusion or ambiguity. Texas’s close proximity to Mexico and its large Spanish-speaking population require the use of bilingual interviewers to overcome language barriers. It is recommended that the entire survey should be completed in approximately two to three minutes (Jessup, Casavant, and Lawson, 2004). A longer

questionnaire may result in impatience and fatigue in the participant, which could compromise the integrity of the collected data.

Figure 5.1 and Table 5.2 provide an example of the proposed truck travel questionnaire that can be used in roadside intercept surveys aimed at collecting truck data for the SAM and the explanation of the data fields, respectively.

Table 5.2 Explanation of the Data Fields

Field Name	Description
Survey location	A unique number should be assigned to each survey site
Date	Survey date
Surveyor ID	A unique number should be assigned to each surveyor
Time	Time survey is conducted
Direction surveyed	Direction in which truck traffic is surveyed
Vehicle classification	Truck and trailer configuration
Company name	Name of trucking company as printed on the driver's door
City	City as printed on the driver's door
State	State as printed on the driver's door
Telephone number	Telephone number on the driver's door
Tractor license plate registration	License plate registration number of the tractor
Trailer license plate registration	License plate registration number of the trailer
Commodity	The major commodity carried
Shipment type	Whether it is a truckload or less-than-truckload shipment*
Payload weight	The weight of the cargo that is carried
Cargo value	The value of the cargo that is carried
Cargo origin city	City where the cargo was picked up
Cargo origin state	State/province where the cargo was picked up
Cargo origin address	Address where the cargo was picked up
Cargo destination city	City where the cargo will be dropped off
Cargo destination state	State where the cargo will be dropped off
Cargo destination address	Address where the cargo will be dropped off

*A truckload shipment weighs either in excess of 10,000 pounds or does not allow a truck operator to carry any other load. A less-than-truckload shipment weighs less than 10,000 pounds and does allow a truck operator to carry other loads on the same truck. Usually package carriers such as FedEx, UPS, and the U.S. Postal Service are not considered less-than-truckload carriers (Federal Highway Administration, 1996).

5.2.9 Interview Procedure

Trucks selected for an interview should be directed to a designated area. As the selected truck is parking, the interviewer should record information, such as the time of day, the truck configuration, and the company information, that is available through visual observation to minimize the amount of time required and inconvenience to the truck driver.

Once the truck is parked, the interviewer should approach the truck driver in a friendly and polite manner and request cooperation in the survey. The driver should be informed that participation in the survey is voluntary. Once a driver agrees to participate, the interviewer should go through the questions as quickly as possible while maintaining proper quality control.

If, for any reason, there is inadequate space for trucks to park safely and an interviewer is not available, a truck driver who would have participated should be allowed to leave (Gillis, 1994). It is thus very important to have an adequate number of surveyors available to collect data.

5.2.10 Data Quality Control

Errors are inevitable when collecting data. Some examples of errors that could occur during data collection include the following (Jessup, Casavant, and Lawson, 2004).

- Systematic errors caused by inappropriate interview procedures, inappropriate site selection, or ambiguous questions
- Inaccurate responses given by the truck drivers
- Inaccurate recording of vehicle data or driver responses by the surveyor

Effective data management will help to reduce errors from mistakes in data collection or database entry. The number of systematic errors that may result from ambiguous questions can be reduced by conducting pilot tests of survey instruments to identify problems with questionnaire wording. Any inappropriately worded or ambiguous questions should be addressed appropriately in updated versions of the survey. In addition, ongoing improvements to survey instruments should be encouraged. Feedback from both the survey crew and supervisors involved in the data collection process should be collected and considered. By improving the clarity of the interview questions, drivers are less likely to provide inaccurate responses, which will further reduce potential errors (Gillis, 1994).

An informative training program is a form of data quality control that is established well before the actual surveying takes place. Training should, however, not end once the surveying begins. Ongoing training and supervision provide an opportunity for answering questions that arise and will remind the survey crew of the interview procedures, which will reduce the number of errors caused by inaccurate data recording. Supervisors should also check completed questionnaires immediately for accuracy and completion. Any problems can thus be addressed immediately.

Finally, logic checks should be performed on all questionnaires before being entered into a database. For example, a potential error that may occur is when the driver provides the gross vehicle weight of both the cargo and the truck instead of the weight of the cargo being transported. Certain assumptions thus need to be established for handling such errors before the data is entered into a database. For example, if the cargo weight provided exceeds the legal limit for that truck's particular axle configuration, the weight will be assumed to be the gross weight of the truck and cargo. Reference questions could also be included in the survey to facilitate the checking of answers that truck drivers provide and team members record.

5.2.11 Data Entry and Clean Up

Capturing data electronically through the use of handheld computers or other devices will reduce data entry errors. As an alternative, data entry forms with numerous built-in checks can be programmed to facilitate the accurate entry of data. For example, the data that can be entered in certain data entry boxes can be restricted. As an example, Question 4 in the survey has thirteen

possible choices, ranging numerically from 1 to 13. The software can thus be coded to display an error message if the data entry person attempts to enter a value greater than 13. Also, entries can be restricted to those included in a drop-down list. In this case, the data entry person will be able to only choose from the list. For example, state or province information can be limited to the U.S. states and Canadian and Mexican provinces that are included in the drop-down list.

Once the data has been entered, it will still be necessary to visually inspect the data for missing or incorrect observations (Clark, Jessup, and Casavant, 2002). Sometimes an interviewer might not be able to obtain all the information needed to complete the survey or might accidentally neglect to complete the survey. Instead of disposing of valuable data, some information can be confidently inferred using responses from other questions on the survey. The importance of asking reference questions is also relevant here (Gillis, 1994). At times, however, a survey may be so incomplete that its data cannot be used. Guidelines should be established to define the minimum amount of information needed from a particular survey for it to be considered useful.

Finally, an effort should be made to reduce the effect of accidentally using data from one particular truck more than once. This may happen if data are collected on the same highway from the same truck at two or more interview sites (Gillis, 1994). One option is to compare the license plate registration numbers of the trailer and truck tractor in a specific geographic area to ensure that duplicate entries do not exist for the same truck.

5.2.12 Data Expansion

Because only a sample of trucks are interviewed to represent the characteristics of the population of trucks transporting cargo on Texas roads, it is necessary to develop statistical weights to expand the sample data to reflect the characteristics of the population. A number of statistical weights, modeled after those used in similar studies conducted by Washington State University, can be used to expand sample data (Jessup, Casavant, and Lawson, 2004).

The objective of the first weight is to expand the sample data at each site for each season to reflect the characteristics of the population for that week, site, and season. The weight will be different for different seasons because the number of trucks sampled and the total number of trucks passing the site will be different for each season. The first weight can be calculated as follows: if 100 trucks are sampled for an interview and a total of 600 trucks pass the site in one week in any particular season, the site weight will be 6.0 for that specific season. This number can then be multiplied by any sample characteristic to provide a statistical estimate of the characteristics of the population. For example, if all the truck drivers reported a total cargo weight of 1,000 tons, the estimate of the total aggregate weight passing the site in one week from all trucks is $1,000 \text{ tons} \times 6.0 = 6,000 \text{ tons}$ each week for that season.

The objective of the second weight is to determine the characteristics of the population across seasons—in other words, to determine the characteristics of the average annual weekly truck trips at a specific site. This is achieved by calculating the weighted average factors across the four seasons and applying this factor to the expanded sample data. For example, assume the expanded trucks surveyed during each of the seasons at a specific site are as given in Table 5.3.

Table 5.3 Seasonal Weight Factor

Season	Total Trucks (Expanded Surveyed Sample)	Seasonal Weight Factor
Spring	2,500	0.26
Summer	3,500	0.36
Fall	2,000	0.20
Winter	1,800	0.18

Source: Adapted from Clark, Jessup, and Casavant, 2003.

The seasonal weight factor is calculated by dividing the individual season totals by the total number of trucks surveyed during the four seasons (i.e., 2,500/9,800) to determine the spring seasonal weight factor. Thus, to determine the average annual weekly truck trips at a specific site, each of the seasonal weight factors are multiplied by the total trucks (i.e., expanded surveyed sample) and summed. These weight factors can subsequently be applied to other characteristics of interest captured in the survey (i.e., number of empty trucks, cargo tonnage, value, etc.) and summed to arrive at the average annual weekly characteristics of truck trips passing a specific site.

Finally, because it is recommended to survey twelve sites in each NTAR in Texas, a weight factor to describe the characteristics of the population of truck movements in a specific NTAR can be calculated. Because the average annual weekly truck trips at each specific site will vary substantially, each site in the NTAR has to be assigned a weight. For example, assume the calculated average annual weekly truck trips at the twelve sites in one of the NTARs are as summarized in Table 5.4.

The seasonal site factor is calculated by dividing the individual site totals by the total number of trucks over the twelve sites (i.e., 3,000/27,015) to determine the Site 1A weight factor. Thus, to determine the average annual weekly truck trips in the specific geographic region, each of the site weight factors are multiplied by the average annual weekly truck trips at each site and summed. These weight factors can subsequently be applied to other characteristics of interest captured in the survey (i.e., number of empty trucks, cargo tonnage, value, etc.) and summed to arrive at the average annual weekly characteristics of truck movements in a specific geographic region.

Table 5.4 Site Specific Weight Factor

Site	Average Annual Weekly Truck Trips	Site Weight Factor
1A	3,000	0.11
1B	4,500	0.17
1C	6,000	0.22
1D	1,300	0.05
1E	1,425	0.05
1F	1,480	0.05
1G	1,100	0.04
1H	1,590	0.06
1I	1,645	0.06
1J	1,700	0.06
1K	950	0.04
1L	450	0.02

5.2.13 Unforeseen Circumstances

Harsh weather conditions, including severe thunderstorms or rain, may affect the quality of the data gathered. Any events that are out of the ordinary, such as construction, automobile collisions, or hazardous material spills, should be well documented (Gillis, 1994). Even where unforeseen circumstances may cease data collection temporarily in the case of extreme weather (e.g., tornadoes) or hazardous conditions (e.g., fuel leak), the interview site weight as defined in Section 5.2.12 can still be calculated. However, because fewer truck drivers will be interviewed, the value of the site weight will be much higher than normal. This will, however, assist in the statistical adjustment of the sample data that would otherwise be missing (Gillis, 1994). Good planning requires that alternate backup data collection days be scheduled in the event that unforeseen circumstances result in inadequate data collection.

5.3 Truck Carrier Participation

It has been recognized that effective partnerships are needed between government and the freight community to ensure adequate planning and funding of transportation infrastructure at the state and local levels. Despite this recognition, the transportation planning community struggles to understand the needs of the freight community, partly owing to the inferior freight data that are available to freight planners (Freight Stakeholders National Network, n.d.).

Enhanced freight mobility through infrastructure improvements is in the interest of both the private sector and the transportation planning community. In addition, reliable freight data can be valuable to the private sector in informing investment decisions relating to equipment utilization, new markets, and business opportunities (Transportation Research Board, 2003). This survey approach is based on the hypothesis that a statistically representative sample of truck companies operating in, from, to, and through Texas can be convinced to share their operational data with

TxDOT. This section of the document discusses the anticipated components of such a data sharing partnership.

5.3.1 Identify and Recruitment of Trucking Companies

A number of data sources exist from which a representative random sample of truck carriers operating on Texas roads can be identified (see Table 5.5 below).

Table 5.5 Potential Sources to Identify Trucking Companies Operating on Texas Roads

Data Source	Number of Trucking Companies	Information Captured	Web Address
Texas Workforce Commission	Local trucking without storage: 1,669 Local trucking with storage: 1,485 Trucking, excluding local: 4,199	<ul style="list-style-type: none"> • Type of trucking company • Employer address (in some cases include mailing address) • Contact person and telephone number • Number of employees 	http://www.texasworkforce.org
TxDOT Motor Carrier Registration Database	37,312	<ul style="list-style-type: none"> • Type of carrier/facility • Customer information (name of business) • Telephone number • Mailing address (include city, state, zip) • Site address (include city, county, state, zip) • Number of vehicles covered under insurance policy 	
USA Data (commercial data source)	100,000 prospects	<ul style="list-style-type: none"> • Business name • Address (city, state, zip code) • Contact name, title • Number of employees (range) • Phone/fax number • Year established 	http://mip.usadata.com/usa/pub/
Texas Vehicle Information and Computer Service, Inc. (commercial data source)		<ul style="list-style-type: none"> • Address of the vehicle owner • Vehicle type • Gross vehicle weight 	http://www.tvics.com
U.S. Census Bureau's Business Register	6 million U.S. employer establishments		http://www.census.gov/econ/overview/se0400.html

The research team obtained access to the TxDOT Motor Carrier Registration Database, which contains the company address and telephone numbers of truck companies registered in Texas. However, it seldom contains the name of a contact person in the company. When a

random sample of these companies was approached, the research team found that company operators or employees answering the telephone in most instances act as gatekeepers. It was thus very difficult (a) to identify the appropriate company representative that can make a decision about entering into a data sharing partnership with TxDOT, and (b) to be transferred to the appropriate person. This type of cold calling was very time consuming and did not provide a satisfactory response.

The research team thus approached a number of trucking companies that have been exposed to transportation planning through their involvement with the North Central Texas Council of Government's Intermodal Freight and Safety Committee (IFS). The membership list of the IFS Committee was obtained, and eight trucking companies were interviewed to determine (a) whether the company would consider participating in a truck data sharing initiative with TxDOT, and (b) what their conditions for participation would be.

All the trucking companies interviewed indicated their willingness to participate in a data sharing arrangement with TxDOT provided that certain conditions were met. The following list reflects the conditions for participation in descending order of the number of times mentioned.

- No information about the company will be included in the aggregate database that is compiled and used by TxDOT.
- The data will not be used for law enforcement or litigation against the company.
- The Texas Motor Transportation Association (TMTA) will be involved to protect the interests of those that participate.
- No severe cost burden will be imposed on the trucking company in compiling the data.
- TxDOT will demonstrate to the trucking companies that the data will be used for a worthwhile purpose.
- No shipper details will be requested.
- The trucking company will have access to the aggregated database compiled by TxDOT.

Only one trucking company indicated that TxDOT would have to compensate the company for the costs of extracting the data and providing it to the agency as a condition of participation.

Given this positive response, the research team was encouraged to believe that an extensive public outreach effort, including a news release, and the involvement of the Texas Motor Transportation Association could result in a statistically representative sample of truck carriers being convinced to provide TxDOT with the data required for the SAM. In addition, resources are available to TxDOT to assist in recruiting trucking companies. One such resource is the Freight Stakeholders National Network. The network can (a) recruit members from their constituencies, (b) provide policy support and technical sources to the group, (c) identify and support needed freight investments, and (d) share examples of best practices from other groups in other states (Freight Stakeholders National Network, n.d.).

5.3.2 Sampling

Sampling Trucking Companies

For data collected from a sample, sample size is a key determinant not only of data quality and reliability, but also of cost. Statistical formulas exist to determine the number of trucking companies that have to participate in a data sharing arrangement with TxDOT to ensure a statistically representative sample (see the text box). It is suggested that a random sample of trucking companies are selected and approached to participate in the data sharing initiative. Once the sampled list of trucking companies has been determined, TxDOT should

involve, among others, the TMTA and Metropolitan Planning Organizations with freight committees to identify the appropriate contact person in each company.

At the same time, a large number of trucking companies in Texas (approximately 41 percent) have only a single truck insured in the TxDOT Motor Carrier Registration Database. It is foreseen that these smaller companies will be particularly challenging to involve in a data sharing relationship, because the owner is frequently the driver of the truck.

Sampling Shipments

Like the approach adopted by the Commodity Flow Survey, it is recommended that the trucking companies are asked to provide TxDOT with data on the value, weight, commodity, origin, destination, and whether it is a truckload or less-than-truckload shipment for all their shipments in an assigned week four times a year. For shipments that include more than one commodity, the trucking company will be asked to report the major commodity in terms of shipment weight.

Determining the Number of Trucking Companies to be Sampled

The following statistical formula can be used to calculate the appropriate sample size:

$$n = \frac{[Z_{\alpha/2}]^2 \times 0.25}{E^2}$$

where

n = number of trucking companies to be included in the sample,

$Z_{\alpha/2}$ = critical value corresponding to a specified confidence level (e.g., 1.96 is the critical value for a 95% confidence level), and

E = margin of error (e.g., 0.05).

“For the 2002 CFS, each establishment was assigned a 1-week reporting period every quarter, for a total of 4 weeks in the calendar year. By assigning different reporting periods to different establishments, the sample covered all 52 weeks of the year.”

—National Research Council of the National Academies, 2003

5.3.3 Data Collection

To minimize the data reporting cost burden imposed on participating trucking companies, electronic reporting options should be explored. The 2002 Economic Census developed survey software that allowed for the importation of businesses data directly from company spreadsheets,

thus reducing the burden on respondents in an effort to encourage participation in the survey (National Research Council of the National Academies, 2003). Electronic reporting would also reduce the cost of data entry for TxDOT and reduce reporting errors.

Because some larger trucking companies already use the Internet to transmit manifest information between the shipper and truck driver, it is assumed that the Internet can also be used to share relevant data with TxDOT. Options to transmit this information securely using Secure Socket Layers (SSL) and encryption are available.

During the discussions with the eight trucking companies interviewed, it became apparent that it would be more difficult for smaller trucking companies to use electronic reporting options because their systems are not as sophisticated. To ensure participation, every effort should be made to ensure that the labor cost burden imposed on trucking companies, especially the smaller companies, in compiling the data is minimized. Most of the trucking companies, however, agreed that submitting the data electronically would enhance their participation. For the larger trucking companies with computerized systems, it will be relatively easy to run queries and provide TxDOT with the data needed, because the requested data are readily available in their systems.

5.3.4 Data Expansion

The data expansion procedure would be similar to that for the roadside intercept surveys. The sample data obtained can be expanded to reflect the truck travel characteristics for the population of trucking companies by using the inverse of the sampling rate. For example, if 10 percent of the trucking companies of a particular size were surveyed, the expansion factor would be 100/10, which equals 10. This first weight is used to expand the sample data for each season to reflect the characteristics of the population for that week and season.

The objective of the second weight is to determine the characteristics of the population across seasons—in other words, to determine the characteristics of the average annual weekly truck trips. This is achieved by calculating the weighted average factors across the four seasons and applying this factor to the expanded sample data (see Section 5.2.12).

5.4 Cost Effectiveness

As was shown in Chapter 2 of this report, most state departments of transportation have purchased the Reebie TRANSEARCH database as the primary source of freight data in their freight planning models and studies. The database is a relatively inexpensive source of detailed freight data at a cost of approximately \$64,000 in 2004 for the data required by the SAM. Several questions have, however, been raised about the robustness of the Reebie TRANSEARCH data, because the sources of the data, the methodology used, and the assumptions made in compiling the database are confidential. It is therefore not possible to answer questions about the validity and reliability of the Reebie TRANSEARCH database.

At the same time, few states rely on the collection of primary freight data, because this can be a costly and time-consuming process at the state level. Washington State's Eastern Washington Intermodal Transportation Study (EWITS) was a six-year, \$1.94 million effort. The statewide origin–destination roadside intercept truck surveys conducted in 1993–1994 were a major component of the EWITS study. The data collected from the 28,000 truck driver interviews at twenty-seven different locations are widely regarded as providing useful freight

movement data for Washington State and statistically reliable information on truck characteristics and commodity flows for all major Washington highways. Texas is, however, a much larger state than Washington and has significantly more road infrastructure, as compared with Washington. For example, Texas has 14,988 lane-miles of interstates, compared with the 3,941 lane-miles of interstates in Washington (Federal Highway Administration, 2002). To ensure geographical coverage of the state of Texas and to account for variations in the characteristics of trucks that use different roadways, it is recommended that a minimum of ninety-six sites be identified: three sites per major highway type (i.e., interstate, state highway, Texas highway, and farm-to-market/ranch-to-market) in each of the eight NTARs of Texas. Also, to account for time-of-day, day-of-week, and seasonal variation, it is recommended that each site be surveyed for 24 hours, seven days a week, in each of the four seasons. Using the external survey bid prices of surveying companies in 2004 as a guideline, the cost of populating the SAM database through roadside intercept surveys would be prohibitive. The total cost of compiling the truck database would be largely a function of the statistical reliability that is required, which is influenced by the degree of geographical coverage and the number of days surveyed. It is estimated that surveying the ninety-six sites only one day per week four times per year would cost TxDOT in excess of \$5 million, using the bid prices by surveying companies for 2004 as a guideline. Some options—for example, using volunteers from the NTARs as surveyors—could result in significant savings in labor, travel, and per diem costs and are worth exploring.

No previous precedent exists for collecting statewide truck travel data through a data sharing partnership with trucking companies. It is thus very difficult to estimate the costs associated with this approach. Initially, a substantial cost component would be the cost of recruiting a statistically representative sample of the trucking companies using Texas roads to participate in the data sharing initiative. The TMTA would have to be involved and compensated, and a significant public outreach effort would be required. There would also be initial costs in determining the most appropriate data collection methods for obtaining data from companies that not only vary in size but also in the level of sophistication in how the required information is captured by the trucking company. There will thus be upfront costs associated with developing electronic reporting options, as well as less sophisticated paper-based options (i.e., questionnaires) and data entry software to facilitate the capturing of the relevant information by TxDOT or its appointed consultant. It is estimated that the costs of recruiting and developing appropriate data reporting options would be significant, but most of these costs would be required only initially in convincing the trucking companies to enter into the partnership with TxDOT and to facilitate the data collection process. After a statistically significant sample of trucking companies have entered into an arrangement to provide TxDOT with the necessary data required for SAM, the recurring costs would be limited to maintaining a dialogue and relationship with the trucking companies involved, the costs associated with compiling the database, data quality management, and the calculation of the necessary expansion factors.

5.5 Concluding Remarks

Collecting primary freight data through the two approaches highlighted in this chapter would be relatively costly, compared with purchasing the Reebie TRANSEARCH database. Surveying ninety-six sites one day per week four times per year would cost TxDOT in excess of

\$5 million, using the bid prices by surveying companies for 2004 as a guideline, although some cost-saving options exist. Because there is no precedent for collecting statewide truck travel data through a data sharing partnership with trucking companies, it is more difficult to estimate the costs associated with this approach. It is, however, foreseeable that this will entail a large upfront cost arising from the recruitment of trucking companies, the development of software to minimize the burden on trucking companies in submitting the data, and finally labor costs associated with compiling and managing the database. Once these companies have entered into such a partnership, the annual costs would be limited to liaising with the companies and compiling and managing the database. These costs should, however, be weighted against the benefits of a reliable database that could be used with confidence for freight planning.

6. Future Opportunities for Collecting State Truck Data

It is anticipated that a number of initiatives that are currently considered at the national level might result in more reliable state-level truck data in the future. The events of September 11, for example, have resulted in a new emphasis on the need for timely and accurate shipment data to address security concerns. Two initiatives, funded by the U.S. government and U.S. Customs and Border Protection, which would enhance the quality of available truck data, are briefly highlighted. The chapter also mentions the national freight data program proposed by a Transportation Research Board committee to address the general lack of freight data available to transportation planners and decision-makers. Finally, the chapter summarizes the data attributes captured by current Intelligent Transportation System (ITS) technologies before highlighting a new federal ITS initiative that could potentially provide state and local transportation planners with more robust truck data.

6.1 Security Initiatives

The American Transportation Research Institute conducted a study, funded by the U.S. government, from the summer of 2000 through December 2002 that had the following objectives (American Transportation Research Institute, 2002).

- The operational testing of a Electronic Supply Chain Manifest (ESCM) designed to *“improve cargo security”* and to identify the personnel involved in cargo transactions
- To increase productivity by *“expediting cargo processing, reducing manifest lead times, and reducing the probability of human error during data entry.”*

The ESCM allowed manufacturers to send cargo information in real-time, using encrypted Internet software, to trucking companies, consolidators, freight forwarders, and airlines in advance of required pick-ups. In Phase II, biometric imprinting technologies (fingerprint recognition) were tested to identify enrolled personnel involved in the cargo transactions. In addition to the security benefits, the ESCM also reduced the number of information errors at different points along the distribution channel. According to the American Transportation Research Institute (2002), “These improvements to the data transfer process as well as the incorporated security steps add value to the truck-to-air cargo transport operation.” The cargo data captured by the manifest include airport of departure, requested routing, weight/value, airport of destination, and so forth. The test involved manufacturers/shippers, motor carriers, and airlines in the Chicago O’Hare and New York City JFK airport areas (American Transportation Research Institute, 2002). This research has contributed to one of the nine new FHWA initiatives, specifically the nationwide deployment of a universal electronic freight manifest to facilitate the movement of freight through security checkpoints and across modes.

U.S. Customs and Border Protection is developing the Automated Manifest System (AMS), which is “a multi-modular cargo inventory control and release notification system for sea, air, and rail carriers” and the Automated Commercial Environment (ACE), which aims to expedite the flow of cargo and entry processing by providing “participants with electronic authorization to

move cargo prior to arrival” (American Transportation Research Institute, 2002). ACE will eventually be integrated with the proposed International Trade Data System (ITDS) that is scheduled for deployment in 2005–2007. The ITDS, a federal government project, will “electronically collect, store and disseminate all international trade data” (American Transportation Research Institute, 2002). It is predicted that the ITDS will be the single point to submit trade data to and receive trade data from (American Transportation Research Institute, 2002). Timely and robust freight data, captured for security reasons, might thus become available to state planners in the future for transportation planning.

6.2 Proposed National Freight Data Program

Conference participants at the Data Needs in the Changing World of Logistics and Freight Transportation conference, hosted by the New York State Department of Transportation (NYSDOT) on November 14–15, 2001 agreed that “currently available regional and national data are inadequate to support the requirements of analysts and policy makers and that market area data are not readily available.” At the same time, conference participants noted that an understanding of why data are needed is required before resources are invested to collect and distribute freight data (Transportation Research Board, 2003). A national freight database was proposed by an appointed ten-member Transportation Research Board committee to address the issue of inadequate freight data. The following variables were proposed for inclusion in the national database (Transportation Research Board, 2003):

- shipment origin and destination,
- commodity characteristics, weight, and value,
- modes of shipment,
- routing and time of day, and
- vehicle or vessel type and configuration.

It is, however, unclear whether such a national database would provide adequate state coverage of truck movements required for statewide planning. Although it is generally agreed that a national freight data framework is a U.S. Department of Transportation priority, there was no timeline for implementing such a program as of the publication date of this report.

6.3 Real Time Truck Data Collection Methods

Current ITS technologies can collect routing, time, carrier, origin, and destination data for truck movements, but not commodity detail or truck characteristics. This section of the report provides an overview of the currently available ITS technologies. Table 6.1 provides a summary of the attributes captured by each of these technologies.

Inductive loop detectors. Rectangular- or square-shaped wire loops buried below the surface of the pavement. Every time a vehicle passes over the loop, the change in electrical inductance registers a vehicle count. Although the capital, installation, and data collection costs associated with loop detector systems tend to be relatively low to moderate, the data accuracy is

also comparatively low. This is largely due to the high failure rates associated with inductive loops, which results in inaccurate estimations (Riley, 1999).

Sensors (acoustic, infrared, and radar/microwave). Acoustic sensors record the sound energy of passing vehicles through microphones that are aimed at the traffic stream. Infrared sensors detect passing vehicles by measuring the time for a reflected signal emitted by a low-energy laser beam at the road surface to return to the device. Radar/microwave sensors detect a passing vehicle through the frequency difference of a transmitted signal (i.e., microwave radiation) and the frequency of the received signal. These sensors can differentiate between moving vehicles and stationary vehicles, determine the speed of passing vehicles, and have a range finding ability that enables them to sense multiple zones. Although they are expensive to install, radar devices do not deteriorate from traffic wear.

Automatic vehicle classification (AVC) recorders. These recorders consist of two sensors (e.g., road tubes, in-pavement inductive loops, piezo axle sensors) spaced apart across the road surface to record the axle spacing of the passing vehicles. In addition, the date and time, number of axles per vehicle, and travel speeds can be recorded. If electronic credential (EC) technology is used together with AVC recorders, an electronic record of the vehicle type and contents can be captured.

Automated vehicle identification (AVI) systems. These systems consist of three components: (a) a transponder or electronic tag attached to the vehicle, (b) a roadside communication unit, and (c) an electronic storage and data processing computer system. Toll tags have unique identification numbers embedded in the tag and transmit this information when requested by a roadside toll tag reader. Toll tags are relatively inexpensive and allow vehicles to pass through toll plazas without having to stop to pay tolls. Commercial vehicle operations (CVO) tags also have a unique identification number embedded in the tag and can store data. The tag can thus transmit the unique identification number and the stored data when requested by a roadside reader. Many large trucking companies use CVO tags in their normal operations. In general, AVI systems require significant investments in capital and installation costs. Data collection costs are, however, relatively low, although limited to fixed routes and checkpoints. The accuracy of the data collected through AVI systems is considered very high (Riley, 1999).

Weigh-in-motion (WIM) equipment. Portable WIM equipment usually consists of two loops and a capacitance weigh pad. Permanent WIM equipment usually consists of inductive loops and one or more axle weight sensors. The most common axle weight sensors are piezoelectric sensors, bending plates, single load cells, and fiber optics. Typically, date and time, vehicle lengths, speeds, axle weights and spacing, and gross vehicle weight can be recorded.

Video image detection. Traffic volumes, speeds, and vehicle classifications are calculated from videotapes recorded with a video camera (i.e., closed-circuit TV cameras) at the survey site. Different techniques (i.e., tripline and tracking) exist to analyze the video image. Tripline detects a vehicle by monitoring specific zones of the video image. Tracking uses algorithms to identify and track passing vehicles on the video image. Typically, the digital clock in the video image is used to determine time intervals.

Automatic vehicle location (AVL)/global positioning systems (GPS). AVL systems use GPS technologies and satellites to record the date, time, and location of a vehicle as it traverses the transportation network at specified time intervals. Many large trucking companies have invested in AVL systems to optimize routing and dispatching schedules, to track and navigate shipments, and to manage their assets. The upfront capital and installation costs associated with

these systems tend to be very high, although the data collection costs are comparatively low. The accuracy of the data collected is, however, considered very high (Riley, 1999).

License plate matching (LPM) systems. LPM systems involve at least two video cameras placed separately along a road segment. The images of the license plates of the downstream vehicles are then matched with the images of the upstream vehicles to determine travel time and speed of the vehicles.

Table 6.1 Attributes Captured by Current ITS Technologies

Attribute	Inductive Loop Detectors	WIM Sensor Systems	Sensors	Video Image Detection	AVC	AVI Toll and CVO Tags	AVL GPS	License Plate Matching
Vehicle classification	X	X	X	X	X	X		
Vehicle weight		X						
Vehicle speed	X	X	X	X		X	X	X
Vehicle delay data	X		X	X				
Vehicle incident data	X		X	X				
Traffic volume data (classification)	X	X	X	X	X	X		
Commodity / cargo type				X	X	X	X	
Payload (cargo) weight								
Truck O-D patterns						X	X	X
Trip O-D patterns						X	X	X
Average tour length						X	X	X
Number of stops per tour						X	X	X
Number of truck stops for LTL shipments						X	X	X
Travel time	X		X			X	X	X
Transit time						X	X	
Travel time reliability						X	X	

Dedicated Short Range Communications

Dedicated short range communications (DSRC) devices are being developed by the FHWA, the American Association of State Highway and Transportation Officials, a number of state departments of transportation, some of the worlds largest electronic companies, toll authorities, car and truck manufacturers, and the American Trucking Association. DSRC is a medium range communication service intended to support both public safety (e.g., collision avoidance) and licensed private operations over roadside-to-vehicle and vehicle-to-vehicle communication channels. It is anticipated that these devices will be installed in all vehicles—cars

and trucks—during the manufacturing process. Each DSRC device will contain a device identification number and has been designed to replace toll tags, CVO tags, and PrePass tags. In addition, it is predicted that these devices could be the underlying technology for electronic waybills that will make the collection of commodity data feasible. Currently, it is anticipated that deployment will commence in 2008.

6.4 Concluding Remarks

This chapter highlighted a number of initiatives that could provide TxDOT planners with better freight data. At a minimum, TxDOT should stay abreast of the initiatives that are currently being developed at the national level. Ideally, however, the agency should participate in the workshops that are currently planned as part of the development of the DSRC devices to ensure that the design of the devices allows for the collection of robust truck data that can be used for state and local transportation planning.

7. Freight Forecasting

Freight forecasting procedures can range from very simple methods to very complex, data-intensive models. In general, statewide freight demand forecasting techniques can be grouped into three different types: causal methods, trend analysis, and qualitative methods (Memmott, 1983). These methods are briefly highlighted in this chapter. In addition, the state-of-the-practice in statewide truck forecasting is discussed before the chapter concludes with recommendations to conduct freight forecasting in the SAM.

7.1 Statewide Freight Demand Forecasting Techniques

7.1.1 Causal Methods

Regression Analysis

Regression models are considered robust for freight forecasting. Regression analysis entails the identification of one or more independent variables that may influence the value of a dependent variable. A mathematical relationship between the dependent and independent variables are subsequently defined to allow the analyst to determine the forecasted change in the dependent variable from forecasted changes in the independent variables. In other words, “regressions normally use historic time-series data (an alternative is cross-section data) obtained for both the dependent and independent variables over the course of several time periods (e.g., years). Regression techniques are applied to the historic data to estimate a relationship between the independent variables and the dependent variables; and this relationship is applied to forecasts of the independent variables for one or more future time periods to produce forecasts of the dependent variable for the corresponding time periods” (Federal Highway Administration, 1996).

To forecast the growth of the dependent variable, forecasted data must be available for all the independent variables. “For freight planning purposes, the dependent variables normally would be some measure of freight activity and the independent variables usually would include one or more measures of economic activity (e.g. employment, population, income)” (Federal Highway Administration, 1996). Socioeconomic data can be obtained from standard federal data sources developed by the Bureau of the Census and the Bureau of Labor Statistics.

Different regression analysis techniques exist, including ordinary least squares (OLS), robust, stepwise, and weighted least squares. OLS is the most prevalent, because each data point is given equal weight in estimating the regression coefficients. Outliers may, however, have a significant impact on the accuracy or reliability of the OLS method of regression analysis. The robust regression technique, on the other hand, uses the OLS regression procedure, but through an iterative process it identifies outliers and minimizes their effects in the model (Brogan et al., 2001). Additional detail is provided on conducting robust regression analysis in Sections 7.3.1 and 7.3.2 of this chapter.

Econometric Modeling

Econometric models are statistically derived from time-series data. These models demonstrate the impacts associated with changing conditions in terms of changes in business volume, employment, income, and population over time. These models attempt to capture interactions within systems that are more complex. In general, econometric models are quite accurate (Memmott, 1983), but they can be very complex and data intensive.

Input–Output Modeling

Input–output (I–O) models, such as those developed by Regional Economic Models Inc. (REMI), provide a quantitative framework to analyze the flows between different sectors in an economy and to determine the impacts of any change in demand on the economy. I–O modeling thus requires the use of economic input indicators (i.e., capital, labor, and land) and output indicators (i.e., production levels and demand for goods and services) to estimate economic activity levels by sector, geographic area, and time, which subsequently must be converted to freight transportation demand estimates (Pendyala, Shankar, and McCullough, 2000). Although these models tend to be very accurate, the data requirements can be prohibitive—at a minimum, detailed data are required at the two-digit Standard Industrial Classification (SIC) level (Memmott, 1983).

7.1.2 Growth Factors

The Federal Highway Administration’s (1996) *Quick Response Freight Manual* included a chapter discussing simple methods to forecast freight demand by applying growth factors to baseline freight or economic data. Two approaches were discussed using growth factors: (1) “based on historical traffic trends” and (2) “based on forecasts of economic activity” (Federal Highway Administration, 1996). These two approaches are explained below.

Based on Historical Traffic Trends

The most simplistic approach to forecast truck travel data is to extrapolate truck traffic flows from historical truck traffic data—the so-called growth factor model—assuming at least two years of data are available. Using available data for two historic years, an annual growth factor can be calculated as follows:

$$\text{Annual Growth Factor (AGF)} = (T2/T1)^{[1/(Y2-Y1)]}$$

where

T1 = freight demand in Year 1, and

T2 = freight demand in Year 2.

The AGF can subsequently be applied to forecast demand (T3) in a future year—for example, Y3—by the following:

$$T3 = T2 \times \text{AGF}^{Y3-Y2}$$

where

T3 = freight demand in future Year 3, and

T2 = freight demand in Year 2.

A numerical example is provided on pp. 3.9–3.11 of the Federal Highway Administration’s (1996) *Quick Response Freight Manual*.

The trend analysis approach is generally not robust, because it directly models the growth in truck traffic and does not consider socioeconomic factors or the underlying behavior of the freight transportation system that impacts the growth in truck flows. The method can, however, be used to generate reasonable forecasts of truck flows for short-term future periods, in which case it can be assumed that the trends in socioeconomic factors remain constant and need not be considered explicitly in the analysis to predict future flows. It is believed, however, that future freight demand can be more adequately explained when the impact of economic factors and activities are considered.

Based on Economic Projections

Future freight demand can be estimated using economic projections of future output given an economic indicator (e.g., employment, population, income, etc.). In other words, if it can be assumed that the demand for transport of a specific commodity is “directly proportional to an economic indicator variable that measures output or demand for the commodity,” then growth factors calculated on the basis of the economic indicator can be applied to freight traffic to forecast future freight demand (Federal Highway Administration, 1996). This procedure requires (a) freight traffic data by commodity type for the base year, (b) base year values for the economic indicator value selected, and (c) forecast year values for the economic indicator value selected. Using this data, the procedure is as follows:

$$AGF = (I_2/I_1)^{1/(Y_2-Y_1)}$$

where

I_1 = the value of the chosen economic indicator for a particular commodity or industry group in Year 1, and

I_2 = the value of the chosen economic indicator for a particular commodity or industry group in Year 2.

This AGF is then applied to the base year traffic for the commodity or industry group to forecast the future traffic for the commodity or industry group.

$$T_f = T_b \times AGF^n$$

where

T_b = the base year traffic for the commodity or industry group, and

n = number of years in the forecast period.

Total freight demand can subsequently be determined by summing the forecasts by commodity or industry groups (Federal Highway Administration, 1996). See pp. 3.11–3.12 of the Federal Highway Administration’s (1996) *Quick Response Freight Manual* for a numerical example. This assumed direct relationship (and identical change) between the change in freight demand and the chosen economic indicator might, however, not be accurate because of changes in the following variables (Federal Highway Administration, 1996):

- value of output per ton
- output per employee
- transportation requirements per ton
- competition from other facilities and modes

7.1.3 Qualitative Methods

Market Research Methods

Market research methods can be used to forecast longer-range developments. They usually involve personal interviews with industry experts, panels, or focus groups in an effort to predict longer-term developments such as major shifts in commodity flows. The Delphi technique, for example, is a structured process that can be used to reach consensus given different perspectives. These methods can be quite accurate (Memmott, 1983) but can also be time consuming and costly.

7.2 State-Of-The-Practice in Statewide Truck Forecasting

Since the passage of ISTEA and subsequently TEA-21, a number of state planning agencies have developed freight demand forecasting models to assist in policy development and investment decisions. This section of the report provides an overview of the state-of-the-practice in truck traffic forecasting.

7.2.1 Freight Analysis Framework (FAF)

The Freight Analysis Framework (FAF) funded by the Federal Highway Administration (FHWA) includes a freight-forecasting component covering forty-eight U.S. states and the District of Columbia. The FAF has been developed using data from the commercially available Reebie TRANSEARCH database. The TRANSEARCH database was used to develop a base year county origin–destination commodity flow matrix, representing annual intercounty commodity tonnage moved by the major freight modes (i.e., truck, rail, air, and water). Future commodity productions and consumptions for each county were computed using the economic growth rates for each industrial sector estimated by WEFA. The forecasted county origin–destination matrix was subsequently developed using a Fratar growth factor model. The Fratar method iteratively computes forecasted intercounty commodity flows as a function of the current flows and the growth factors of the production and attraction zones. The most important limitation associated with the Fratar method is that no impedance measure (e.g., travel costs) is considered in distributing the forecasted commodity generations. Thus, the model does not account for any future changes in the freight transportation network, and the implicit assumption is thus that the network configurations remain the same as those for the base year.

7.2.2 Texas Statewide Analysis Model (SAM) Freight Forecasting Methodology

The freight forecasting methodology embedded in the Texas SAM forecast 2025 freight flows for eleven commodity groups for internal–internal, internal–external, external–internal, and through trips. The forecasting approach is founded in the four-step sequential forecasting

process comprising freight flow generation, distribution, mode split, and assignment. Freight flow generation equations were developed for each commodity group as a function of employment and dummy variables that represent freight activity associated with distribution centers, intermodal transfer facilities, and any other freight generating facilities. Finally, because an increasing trend in worker productivity is observed over time, a worker productivity adjustment factor of 1.55 was applied to the forecast commodity flow estimates produced. Because no reliable relationship between agriculture and raw material commodity flows and employment variables could be established, forecasted flows for these commodities were based on the 2010 to 2020 Wharton Economic Forecasting Associates (WEFA) forecasts obtained by TxDOT. A 2025 WEFA forecast was estimated by extrapolating from 2010 through 2020. The 2025/1998 WEFA commodity growth factors were subsequently used to grow Reebie base year (1998) flows (e.g., in the case of intrastate agriculture and raw material freight flows) or to constrain the SAM forecasting model estimates (e.g., in the case of intrastate chemicals and petroleum, building materials, and miscellaneous mixed freight flows). Through freight flows were forecasted by applying the 2025/1998 WEFA through flow growth factors to the base year Reebie through flows for each commodity group. Finally, future freight flows for Texas-to-Mexico, Mexico-to-Texas, other U.S. states-to-Mexico, and Mexico-to-other U.S. states were based on the Latin American Trade and Transportation Study (LATTs) growth rates (TxDOT, 2003).

7.2.3 Virginia Intermodal Freight Planning Model

The Virginia Transportation Research Council (VTRC; a cooperative organization funded by the Virginia Department of Transportation and the University of Virginia) developed the Virginia Statewide Intermodal Freight Planning Methodology in 1998.

The freight-forecasting component of the methodology started by identifying the key commodities moved in the state of Virginia. Because different commodities have an inherent tendency to be moved by a certain mode (e.g., trucks normally move high-value/low-weight commodities and rail normally moves low-value/high-weight commodities), the commodity classification procedure considered both weight and value to avoid modal bias in identifying key commodities. In total, fifteen key commodities were identified, which accounted for more than 68 percent of the total weight and 52 percent of the total value of all freight shipments in Virginia (Brogan, Brich, and Demetsky, 2001). Secondly, base year county commodity origin–destination flow matrices were developed for the key commodities from the Reebie TRANSEARCH database. Only movements for counties within Virginia were considered. Through movements (i.e., external–external commodity flows) were thus not considered. Thirdly, to generate estimates of future county productions and attractions for each of the key commodities, regression equations were developed to represent commodity productions and attractions in each county as a function of a set of explanatory socioeconomic variables. Robust regression analysis was used to minimize the effect of outliers on the estimates of the regression parameters (Brogan, Brich, and Demetsky, 2001). The explanatory variables considered and the parameter estimates for each of the fifteen key commodities are described in detail by Brogan, Brich, and Demetsky (2001). Future commodity productions and attractions in each county were calculated by using forecasted estimates of the socioeconomic variables for each county from the Census Bureau and the Virginia Employment Commission as inputs into the calibrated

regression equations. Finally, the Fratar growth factor model was applied to generate the forecasted county origin–destination matrices for each of the fifteen key commodities.

7.2.4 Wisconsin Freight Model

The Wisconsin statewide multimodal freight-forecasting model was developed by Wilbur Smith Associates in 1996 as part of Wisconsin’s long-range multimodal transportation plan, called *Translink 21*. The Wisconsin model forecasts flows for thirty-nine commodity groups by truck, rail, air, and water modes between seventy-two Wisconsin counties, thirty-four counties in adjacent states, and thirty-four Bureau of Economic Analysis (BEA) regions representing other states (Federal Highway Administration, 1996). Base year data were developed using the commercially available Reebie Associates’ TRANSEARCH freight data at the BEA level of detail and the 1993 CFS data.

Forecasts were developed for 2020 and five intermediate years using econometric factors derived through trend analysis. “The trendline logic applied consisted of: (a) a change in employment due to a change in production yields a similar change in output, (b) output results in shipments, and (c) commodities can be related to output of a particular industry through SIC codes” (Federal Highway Administration, 1996). Employment was forecasted using economic indicators of industry and the Regional Projection of the BEA. Productivity forecasts were developed from information provided by REMI. Industry output per employee factors were developed to capture the effect of employment change on output. Base year commodity origin–destination tonnage was subsequently multiplied with the “combined ratio of change for employment and productivity to that origin and relevant industry” (Federal Highway Administration, 1996). In the case of certain commodities (e.g., farm outputs, fuels, waste, and nonmetallic minerals), adjustments were made to reflect perceived faster or slower growth than average for all commodities (Federal Highway Administration, 1996).

7.2.5 Indiana Commodity Model

The Indiana Commodity Model was developed by the Transportation Research Center at the Indiana University in 1997 (Black, 1997). This multimodal freight model has the capability to forecast interzonal truck and rail flows for Indiana’s top twenty-one commodity groups among ninety-two Indiana counties, forty-seven contiguous states, and the District of Columbia.

The Indiana freight model follows the four-step travel demand modeling procedure involving freight flow generation, distribution, modal split, and traffic assignment. In the freight flow generation step, OLS regression equations were developed for commodity productions and attractions from the 1993 CFS data as functions of employment and population data obtained from the Bureau of the Census. Future county productions and attractions for each commodity group were estimated from the regression equations using county employment and population projections obtained commercially from Woods and Poole as inputs (Federal Highway Administration, 1999). The forecasted county productions and attractions were subsequently distributed using a constrained gravity model calibrated from the base year CFS commodity flow data (see Section 3.2.3 for a general introduction to the gravity model).

7.2.6 Florida Intermodal Statewide Highway Freight Model

The Florida Intermodal Statewide Highway Freight Model (FISHFM) is in many aspects similar to the Indiana Commodity Model. The FISHFM, however, adopted the modal network and the Traffic Analysis Zone (TAZ) structure from its statewide passenger model. The TAZ structure of the FISHFM thus comprises 508 internal TAZs in Florida, along with thirty-two regional zones for areas outside Florida. The model forecasts freight flows for fourteen commodity groups, including the top ten commodity groups identified from the commercial Reebie TRANSEARCH database.

In the trip generation step, regression equations were developed for the commodity productions and attractions in each zone as a function of zonal employment and population. Future zonal commodity productions and attractions were estimated using employment and population projections from the University of Florida's Bureau of Economic and Business Research as inputs into the calibrated regression equations. Forecasted interzonal commodity flows were developed from the future commodity productions and attractions by applying a gravity model calibrated using base year commodity origin destination data from TRANSEARCH. Modal split analysis was conducted using an incremental logit mode choice model (Cambridge Systematics, 2003). Interzonal truck trips were estimated using payload factors by commodity group and shipment distance class obtained from the Vehicle Inventory and Use Survey (VIUS). These forecasted truck flows were assigned on the highway network using an all-or-nothing traffic assignment approach without considering network congestion.

A unique feature of the FISHFM is the consideration of ports, airports, and rail terminals as special generators of freight traffic. The model designates the fourteen deepwater ports, nineteen airports, and twenty-five major rail terminals in Florida as separate zones for the following reasons: (a) These locations serve as both origins and destinations of nontruck freight movements, and (b) most of the freight flows between these intermodal terminals occur by truck (Cambridge Systematics Inc., 2003).

7.2.7 Mississippi Intermodal Freight Planning Model

The Mississippi Intermodal Freight Planning Model was developed in 2003 by the Mississippi Department of Transportation in cooperation with the Mississippi State University. The model estimates future freight flows for county-to-county, county-to-state, state-to-county, and through commodity movements using data obtained from the publicly available 1997 CFS for the base year. The zones included in the model are the eighty-two Mississippi counties, all states except Mississippi, and the four neighboring states of Alabama, Arkansas, Tennessee, and Louisiana.

Forecasted county-to-county, county-to-state, and state-to-county commodity flows were estimated by applying employment and population growth factors obtained from the Complete Economic and Demographic Data Source (CEDDS; distributed by Woods and Pool Economics, Inc.) to the base year commodity productions and attractions, respectively. Because employment and population growth factors for estimating future county-to-state and state-to-county flows were applied directly to the base year county-to-state and state-to-county origin-destination matrices, freight flow distribution was required only for the county-to-county flows. Future county commodity productions and attractions were distributed among counties using a gravity model. Base year traffic through Mississippi was generated by assigning U.S. state-to-state

flows, excluding the state of Mississippi, to the Mississippi highway network. Consequently, estimating future Mississippi through flows would have required forecasted U.S. origin–destination tables, which was considered outside the scope of the model. On the basis of a rough estimate, through freight flows in Mississippi were forecasted to increase by around 50 percent in the future.

7.3 Proposed Truck Traffic Forecasting Procedure for The Statewide Analysis Model

On the basis of the research team’s review of the state-of-the-practice in statewide truck forecasting, a detailed review of the current truck forecasting method embedded in the SAM, and the team’s experience in the calibration of the MLN models discussed in Chapter 4, the following two methods are proposed as relatively low-cost methods for generating future county-level truck flows for the SAM.

7.3.1 Improve Regression Equations Developed for SAM

The current regression equations used for truck forecasting in the SAM can be made more robust by the following actions:

- incorporating additional explanatory variables in modeling the commodity productions and attractions,
- performing robust regression analysis for calibrating the freight generation equations, and
- developing regression models to represent agriculture and raw material productions and attractions.

Additional Explanatory Variables

The freight generation equations used in the SAM estimate commodity county productions and attractions as a function of county employment and dummy variables to account for freight distribution centers and intermodal transfer facilities. Important socioeconomic variables that could have a significant impact on freight flows, such as county population and income, could be considered in estimating county productions and attractions to improve the goodness-of-fit of the regression models. Forecasts of these variables are available and can be used in the calibrated regression models to estimate future county productions and attractions. These forecasted commodity productions and attractions can be subsequently distributed among counties using an appropriate distribution model.

Robust Regression Analysis

The freight generation equations have been developed by performing nonrobust OLS regression analysis of the dependent variable (i.e., commodity productions/attractions) with the independent variables (i.e., employment and the dummy variables). Freight data are known to contain many outliers owing to sampling and nonsampling errors. Robust regression analysis can thus be used to minimize the effects of outliers on the parameter estimates.

During robust regression analysis, outliers are identified in the data and robust weights are assigned to the observations according to the magnitude of the residuals (robust weight < 0.10 for outliers). High weights are given to observations with small residual values. Observations with large residuals are subsequently eliminated at each stage of a weighted stepwise regression procedure until convergence is achieved on a robust regression model. Thus, observations with smaller residual values have a greater impact on the estimates of the regression coefficients (see the text box in Section 7.3.2 for an explanation of the steps involved in robust regression analysis).

Agriculture and Raw Material Regression Equations

The current forecasting approach does not include freight generation equations for agricultural and raw material commodities, because no reliable relationship could be established between these commodity attractions/productions and the independent variables (i.e., county employment and dummy variables for freight distribution centers and intermodal transfer facilities). Evidently, a linear regression model representing county productions/attractions of agricultural commodities as a function of county employment would have a low goodness-of-fit (low adjusted R^2) because only a small proportion of the variance of the dependent variable (county productions of agricultural commodities) could be explained by the regression model.

When modeling freight flow generations of agricultural and raw material commodities, additional explanatory variables could be considered (e.g., agricultural land area, agricultural sales, agricultural employment⁸, per capita income, employment in the mining sector, and employment in secondary industries that use agricultural and raw material products in the manufacturing of finished products) for inclusion in the regression models in an attempt to capture more of the variance of the dependent variable (productions/attractions of agricultural and raw material commodities) and to achieve a better goodness-of-fit.

7.3.2 Apply Calibrated MNL Models to State-to-State Truck Flow Forecasts

The MNL models discussed in Chapter 4 for estimating county-level truck data in Texas from the publicly available CFS data can be applied to generate future truck flows at the county level. Because the MNL models are essentially truck flow distribution models that estimate Texas county truck flows from state-to-state truck flows, future state-to-state truck flows are required in this approach. This section of the report discusses the steps involved in forecasting truck flows using the calibrated MNL models developed as part of this research project. The steps are as follows.

- Develop robust freight generation equations to model state productions and attractions for each commodity group using base year commodity flow, socioeconomic, and demographic data.
- Generate future state commodity productions and attractions for each commodity group, using the calibrated freight generation equations.

⁸ Data for the agricultural variables suggested are available to the public free of charge from the National Agricultural Statistics Service (see <http://www.nass.usda.gov/census/census02/profiles/tx/index.htm>).

- Compile a forecasted state-level commodity origin–destination matrix from the forecasted state commodity productions and attractions (developed in the previous step) and by applying the calibrated production and attraction flow distribution models developed as part of this research.
- Estimate forecasted state-to-Texas county flows from the forecasted state-to-Texas truck flows and by applying the calibrated MNL production distribution flow model.
- Estimate forecasted Texas county-to-state flows from the forecasted Texas-to-state truck flows and by applying the calibrated MNL attraction flow distribution model.
- Estimate Texas county-to-county flows from the forecasted Texas-to-Texas truck flows and the calibrated MNL attraction and production flow distribution models.

Freight Generation Analysis

The first step in the proposed forecasting procedure is to develop robust freight generation equations to model state productions and attractions for each commodity group as a function of state socioeconomic characteristics. The robustness of the freight generation analysis depends on the selection of appropriate socioeconomic explanatory variables to model the state commodity productions and attractions. The explanatory variables selected should take account of the following (ITE: Trip Generation Handbook, 1997).

- **Measurable.** The explanatory variables should be quantitative measures of the socioeconomic characteristics that impact commodity productions and attractions in the states.
- **Independent.** The explanatory variables should be independent of each other (uncorrelated) to avoid the problem of multicollinearity in the regression models. Collinearity can result in a failure to estimate the regression coefficients or instability in the coefficient estimates (i.e., high standard error in the estimates of the regression coefficients). The text box below highlights some observations that can be used for identifying multicollinearity.
- **State-level data.** Data on the explanatory variables should be available for all U.S. states for the calibration and application of the regression equations to generate state productions and attractions.
- **Reliable forecasts.** Because future state commodity productions and attractions will be estimated using the regression equations, reliable forecasts for the explanatory variables should be available for all the states.
- **Availability of data,** because the truck forecasting approach will be used by TxDOT in the SAM, current and forecasted data for the explanatory variables should be readily available, preferably from public data sources.

Identifying Multicollinearity

Observations that can indicate multicollinearity include the following.

- A drastic change of the regression coefficients when adding or removing a variable.
- A regression coefficient has a negative (positive) sign when it is expected to be positive (negative).
- The explanatory variables have insignificant t statistics, but the regression model has a significant fit based on the F test (analysis of variance).
- An explanatory variable has an insignificant t statistic when it is expected to have a significant impact on the dependent variable.

When multicollinearity is observed, the most practical and frequently used solution is to remove the explanatory variable suspected of causing the problem from the regression model (Greene, 2000).

The Virginia Intermodal Freight Planning model has a robust freight generation component. On the basis of the Virginia freight model, Table 7.1 contains the explanatory variables and data sources that should be considered to model state productions and attractions of commodities.

Table 7.1 Explanatory Variables to Model Commodity Productions and Attractions

Explanatory Variable	Base Year Data Source	Forecasted Data Source
Industry employment	IMPLAN Database	State Department of Labor / Workforce/Employment Commission
Motor freight and warehousing employment	IMPLAN Database	State Department of Labor / Workforce/Employment Commission
Transportation services employment	IMPLAN Database	State Department of Labor / Workforce/Employment Commission
Population	U.S. Bureau of the Census	U.S. Bureau of the Census
Population density	U.S. Bureau of the Census	U.S. Bureau of the Census
State size	U.S. Bureau of the Census	U.S. Bureau of the Census
Per capita income	U.S. Bureau of the Census	U.S. Bureau of the Census
Dummy variables representing freight distribution centers and intermodal transfer facilities	—	—

To reduce the effect of outliers, it is proposed that robust regression analysis (see the text box below) is performed to calibrate state commodity productions and attractions as functions of the explanatory variables included in Table 7.1. Dummy variables for freight distribution centers and intermodal terminals should be included to account for the additional freight activity generated by these facilities.

Stepwise Procedure for Conducting Robust Regression Analysis

The steps involved in conducting robust regression analysis for calibrating state production and attraction equations can be described as follows.

- Perform OLS regression of state commodity productions (attractions) as a function of the explanatory variables.
- Compute the residuals (e) for each state observation:
 $e = \text{observed commodity production} - \text{predicted commodity production from regression equation}$
- Given the residual values, assign robust weights between 0 (for high-value residuals) and 1 (for low-value residuals) to each observation. The following formula can be used to calculate the robust weight assigned to each observation given the residual values (e):

$$\text{Robust weight} = \frac{1}{1 + \frac{|e|}{500}}$$

where $|e| = e$ if $e > 0$
 $= -e$ if $e < 0$.

- Perform OLS regression of state commodity productions (attractions) as a function of the explanatory variables with the robust weights assigned to each observation as calculated in the previous step.
- Compute the residuals for each observation using the newly estimated regression model.
- Re-assign robust weights to each observation on the basis of the magnitude of the residuals.
- Perform the weighted regression iteratively by computing residuals in each iteration and assigning robust weights to observations until the residual values for observations do not change between successive iterations.
- The magnitudes of the robust weights assigned to the observations in the final iteration are used to identify outliers in the data set. Typically, observations with robust weights of less than 0.10 are considered outliers.
- Remove the observations identified as outliers from the data set and perform weighted regression analysis on the remaining observations until convergence to arrive at the final robust regression model for commodity productions (attractions).

The regression equation for state commodity productions as a function of the explanatory variables can be given by the following:

$$P_i = \beta_0 + \beta_1 (\text{industrial employment})_i + \beta_2 (\text{motor freight and warehousing employment})_i + \beta_3 (\text{transportation services employment})_i + \beta_4 (\text{population})_i + \beta_5 (\text{population density})_i + \beta_6 (\text{state size})_i + \beta_7 (\text{per capita income})_i + \beta_8 (\text{DUMMYVAR})_i$$

where

P_i = commodity productions in state i ($i = 1, 2, \dots, 50$),

β_i = regression coefficients to be estimated ($i = 0-8$), and

DUMMYVAR = dummy variable representing freight distribution centers and intermodal transfer facilities.

The regression equation for state commodity attractions as a function of the explanatory variables can be given by the following:

$$A_i = \alpha_0 + \alpha_1 (\text{industrial employment})_i + \alpha_2 (\text{motor freight and warehousing employment})_i + \alpha_3 (\text{transportation services employment})_i + \alpha_4 (\text{population})_i + \alpha_5 (\text{population density})_i + \alpha_6 (\text{state size})_i + \alpha_7 (\text{per capita income})_i + \alpha_8 (\text{DUMMYVAR})_i$$

where

A_i = commodity attractions in state i ($i = 1, 2, \dots, 50$),

α_i = regression coefficients to be estimated ($i = 0-8$), and

DUMMYVAR = dummy variable representing freight distribution centers and intermodal transfer facilities.

State Commodity Production and Attraction Forecasts

Upon calibrating the regression equations for state commodity productions and attractions, future state commodity productions and attractions can be obtained using the forecasted socioeconomic data, highlighted in Table 7.1. Similar to the approach followed in the current SAM forecasting procedure, the dummy variables for freight distribution centers and intermodal transfer facilities will initially be assigned the value 0 (TxDOT, 2003). Given this 0 value, the forecasted commodity productions (attractions) will be computed for each state. The average ratio of forecasted to base year commodity productions (attractions) can then be applied to the original dummy variables of the calibrated regression equations and the state commodity productions (attractions) can be re-estimated for the future.

Compile Forecasted State-to-State Commodity Origin–Destination Matrix

The next step is the development of the forecasted state-to-state commodity origin–destination matrix using the forecasted state commodity productions and attractions as inputs into the MNL production flow distribution model developed in Chapter 4. In other words, the forecasted commodity productions in each state can be distributed to each of the attraction states considering the interstate centroidal distances and the relative attraction levels of the attraction states using the production flow distribution model (see Chapter 4).

$$T_{ij}^k = P_i^k * \frac{e^{U_{ij}^k}}{\sum_{j=1}^{49} e^{U_{ij}^k}}$$

where

T_{ij}^k = forecasted flows of commodity k from state i to state j ,

P_i^k = forecasted productions of commodity k in state i , and

U_{ij}^k = utility for fractional flows of commodity k from state i to state j .

Compile Forecasted County-Level Truck Flow Data

The final step is to compile forecasted county-level truck flow data by applying the MNL models developed in Chapter 4 to the forecasted Texas-to-Texas flows, Texas-to-state flows, and state-to-Texas flows obtained from the forecasted state-to-state commodity origin–destination matrix developed in the previous step. The development of the MNL models and the approach for estimating county-level data from the state-to-state commodity origin–destination matrix are described in detail in Chapter 4.

7.3.3 Conversion of Commodity Flows to Truck Trips

Forecasted county-level commodity flow data ultimately needs to be converted to number of truck trips. It is recommended that the average truck payload factors calculated during the Washington State EWITS study be used to convert commodity flows into commodity truck trips for each of the nine commodity groups (see Table 7.2).

Table 7.2 Average Truck Payload Factors by Commodity Group

Commodity Group	Truck Payload Factor (Tons/Truck)
Agriculture	21.15
Food	17.70
Building materials	16.36
Raw materials	24.08
Chemicals/petroleum	22.03
Wood	16.61
Textiles	11.45
Machinery	11.93
Miscellaneous	17.28

Source: Gillis, Jessup, and Casavant, 1995

Daily truck trips can be computed by dividing the total annual truck trips by the total working days per year. In general, 312 working days are used in the latter calculation to represent a six-day workweek for the trucking industry.

7.4 Concluding Remarks

This chapter provided an overview of six approaches that were reviewed by the research team to determine the state-of-the-practice in truck traffic forecasting. The approaches varied significantly in terms of both the techniques and data sources used, but it was found that most states (a) used the Reebie TRANSEARCH data to develop base year county commodity origin–destination flow matrices, (b) developed regression equations to represent commodity productions and attractions in each county as a function of a set of explanatory socioeconomic

variables, and (c) calculated future commodity productions and attractions in each county using forecasted estimates of the socioeconomic variables from available data sources. Finally, the chapter highlighted two approaches for forecasting truck data for the SAM. The first approach attempts to enhance the current regression equations used for truck forecasting in the SAM by (a) incorporating additional explanatory variables in modeling the commodity productions and attractions, (b) performing robust regression analysis for calibrating the freight generation equations, and (c) developing regression models to represent agriculture and raw material productions and attractions. The second approach forecasts truck flows using the calibrated MNL models developed as part of this research project. The relevance and robustness of the recommended approaches can be determined only once the regression equations have been calibrated. Both these approaches will, however, take advantage of the resources already invested by TxDOT in developing a forecasting approach for the SAM and the current research project.

8. Conclusions and Recommendations

This report documented research that has culminated in the recommendation of various approaches available to Texas Department of Transportation (TxDOT) planners for collecting and maintaining county-level truck travel data in a format that is required for the Statewide Analysis Model (SAM), the model developed to provide a regional model for the state of Texas that focuses on intercity passenger and freight travel.

8.1 Freight Data Used by State Departments of Transportation

State departments of transportation require disaggregated truck flow data to determine the impact of truck movements on the state's road infrastructure—bridges and pavements—and the implications in terms of funding. In addition, robust truck data enable state transportation planners to (a) evaluate strategies for improving freight mobility, (b) forecast system performance, (c) mitigate the impacts of truck traffic, (d) determine the impacts on air quality, and (e) improve the safety and security performance of the road network. Informed decisions, however, are critically contingent upon the availability of accurate data and sound models. Chapter 2 reviewed the freight data used by state departments of transportation in their freight planning models and studies. It was found that state departments of transportation (a) rely predominantly on traffic and classification counts conducted, (b) on the data compiled and published by federal agencies, (c) on the Reebe TRANSEARCH freight database, or (c) to a lesser extent on the collection of original data. The collection of primary freight data at the state level tends to be costly and time-consuming. The most comprehensive statewide origin–destination truck intercept surveys of truck drivers were conducted in Washington in 1993–1994. A few states (e.g., Mississippi) have developed procedures to disaggregate truck flows from publicly available freight data sources, such as the CFS, and available county socioeconomic data gathered by the U.S. Bureau of the Census (e.g., county business patterns and county population estimates).

8.2 Available Freight Data Sources

The research team also reviewed more than fifty private and public freight data sources as part of an extensive literature review (see 4713-P2 entitled *State-Of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S.* for additional details). It was found that only six of the reviewed databases capture some of the variables of interest to TxDOT—namely, truck volume (tonnage, value, and number of loads), origins and destinations, commodities, truckload or less-than-truckload—at various levels of detail. These were (a) the CFS database, (b) the Reebe Associates' TRANSEARCH database, (c) the Freight Transportation and Logistics Service data, (d) the Transborder Surface Freight database, (e) the North American Trucking Survey, and (f) the Vehicle Inventory and Use Survey. These databases differ substantially in terms of the variables captured, scope, and structure. Only the Reebe TRANSEARCH database contains all the variables necessary to populate an intercity and interstate truck travel database required for the SAM. In addition, a number of these data sources were found to be incompatible, largely for the following reasons.

- The detail reported in terms of sample sizes, assumptions, reliability, and so forth prevented comparison of the different sources.
- The protocols for assigning the origins and destinations of truck traffic differs. For example, uncertainty exists as to whether the assigned origins and destinations are the origins of production and destinations of consumption or the truck trip origins and destinations. There is also the issue of whether the origin is the origin of the shipment or the listed address of the shipper and confusion about whether the origin or destination is the administrative port of entry/exit, where the paperwork is filed, or the physical port of entry/exit, where the shipment physically enters or exits the U.S.
- The commodity classifications used are different.
- Different expansion factors and control totals are used.

Given these incompatibilities, the research team concluded that the eventual quality and reliability of a database that results from the “fusion” of different databases will be compromised.

8.3 Populating State Transportation Models Using Public Data

Chapter 3 highlighted the data requirements of the freight component of the SAM. The chapter also reviewed the robust regression and gravity models that states have used to generate county-level truck flows from available commodity flow and socioeconomic data. The gravity model distributes estimated truck flows as a function of the production/attraction levels at the origins and destinations and an impedance measure. The model is thus limited in its consideration of other direct and indirect factors that can influence freight flow distributions. Calibration of the gravity model also requires successive iterations, so parameter estimation can become cumbersome. The research team explored the use of a multinomial logit (MNL) model to estimate county-level truck data for Texas from the CFS database.

8.4 Proposed Approach for Estimating State Truck Data

Chapter 4 summarized the proposed MNL approach to estimate county-level truck travel data from the publicly available CFS and IMPLAN data. MNL models are first calibrated at the state level and then used to estimate truck flows at the county level. Two state-level MNL models were developed for each commodity category:

- The MNL production flow distribution model estimates the fraction of the total productions in a state moving to each attraction state by truck on the basis of the attributes of the attraction states and interstate centroidal distance that serves as a proxy for the generalized cost of transportation.
- The MNL attraction flow distribution model estimates the fraction of the total attractions in a state originating from each of the production states by truck on the basis of the relative production levels of the origin states and the interstate centroidal distance that serves as a proxy for the generalized cost of transportation.

The calibrated state-level MNL production and attraction flow distribution models are then used to estimate Texas county-to-county, state-to-Texas county, and Texas county-to-state truck flows. The modeling steps involved in estimating the county-level truck data for Texas was discussed in detail in this chapter. The suggested approach has the following advantages.

- The MNL production and attraction truck flow distribution models can be calibrated relatively easily using the publicly available CFS truck and IMPLAN data.
- The approach is relatively cost effective, because the models were calibrated using publicly available CFS truck data.
- The degree to which the model predicts the observed truck distribution, measured by the adjusted R^2 , attests to the accuracy and the reliability of the model. The adjusted R^2 values were found to be between 0.4 and 0.5 for most of the commodity groups, which are reasonable considering the limitations of the data used.
- Because the inherent assumptions, limitations, and margins of error associated with the CFS data are well documented, the accuracy and reliability of the data generated from the model can be determined.

Although MNL models are particularly suited for modeling the distribution of truck flows and can be calibrated relatively easily, their accuracy is impacted by the quality and reliability of the data used and the predictive power of the explanatory variables. As such, the approach suffers from the limitations of the CFS data (e.g., unreported data, industry coverage, and a lack of routing details to estimate through truck flows in Texas) and a general lack of data that resulted in the assumption that the distribution of truck flows is a function of the centroidal distances and attraction (production levels) of the states and Texas counties. A number of factors other than distance and the production/attraction levels of states can impact the truck flow distribution between states, including (a) long-standing trading relationships, (b) the location of intermodal terminals, inland ports, and seaports, and (c) a number of mode choice factors. Because these factors (explanatory variables) have not been considered in the model, it affects the accuracy of the results.

Finally, because one of the primary data sources in compiling the Reebie TRANSEARCH database is the CFS data, the research team compared the model estimates with the TRANSEARCH data for Texas. The paired sample t test was used to determine whether there was any statistically significant difference between the model estimates and the Reebie data used in the SAM. The paired sample t test is used to compare the means of two dependent set of samples. In comparing the means of the two data sets, comparisons were made for origin–destination truck flows that are reported in both the databases. The paired sample t test thus determines the differences in the paired data and reports the probability that the mean of the paired differences is zero, given a specified confidence interval. A confidence interval of 95 percent was specified. For six of the nine commodity groups, the mean of the model estimates for Texas county-to-state truck flows was statistically similar to the mean of the TRANSEARCH

data for these flows. In the case of state-to-Texas county truck flows, the model estimates were statistically similar to the TRANSEARCH data for four of the nine major commodity groups, and for Texas county-to-county truck flows, the model estimates were statistically similar to the TRANSEARCH data for five of the nine commodity groups. It is thus believed that the MNL approach, despite the limitations highlighted, provides a cost-effective alternative for obtaining county-level truck flows for Texas for at least some of the commodity groups in the short term.

8.5 Truck Travel Survey Methods

Primary freight data collection involves collecting freight flow data directly from the freight community (i.e., shippers, carriers, receivers, and freight forwarders) through surveys, including roadside intercept surveys, mail-out/mail-back questionnaires, combined telephone-mail-out/mail-back questionnaires, or telephone interviews. Done correctly, these survey methods are, in general, the most reliable and accurate methods of obtaining freight flow data for statewide freight planning programs. Collecting primary freight data is, however, a costly and time-consuming process, especially when conducted at the state level. Chapter 5 provided an overview of the available primary freight data collection methods that have been used and discussed two data collection approaches—an extensive program of truck intercept surveys and truck carrier participation—that showed the most promise of providing TxDOT with the data needed for the SAM over the medium term. The more costly of the two approaches would be the collection of statewide truck data by means of roadside intercept surveys. The total cost of compiling the truck database would be largely a function of the statistical reliability that is required, which is influenced by the degree of geographical coverage and the number of days surveyed. To ensure geographical coverage of the state of Texas and to account for variations in the characteristics of trucks that use different roadways, it is recommended that a minimum of ninety-six sites be identified: three sites per major highway type (i.e., interstate, state highway, Texas highway, and farm-to-market/ranch-to-market) in each of the eight National Transportation Analysis Regions (NTARs) of Texas. To account for time-of-day, day-of-week, and seasonal variation it would be ideal if each site were surveyed for 24 hours, seven days a week in each of the four seasons, but the costs would be prohibitive. Even surveying the ninety-six sites only one day per week four times per year would cost TxDOT in excess of \$5 million using the bid prices by surveying companies for 2004 as a guideline. Some options, however, exist to reduce the surveying costs and are worth exploring—for example, using volunteers from the areas as surveyors could result in significant savings in labor, travel, and per diem costs.

The truck carrier participation approach is based on the hypothesis that a statistically representative sample of truck companies operating in, from, to, and through Texas can be convinced to share some of their operational data with TxDOT. The research team contacted eight trucking companies that have been exposed to transportation planning through their involvement with the North Central Texas Council of Government's Intermodal Freight and Safety Committee to determine whether these companies would consider participating in a data sharing initiative with TxDOT and what their conditions for participation would be. All eight representatives indicated their willingness to participate in a data sharing arrangement with TxDOT provided that certain conditions can be met. The most often mentioned conditions were that (a) no information of the company will be included in the database compiled and used by TxDOT, (b) the data will not be used for law enforcement or litigation against the company, (c) the Texas Motor Transportation Association (TMTA) will be involved to protect the interests of

those that participate, and (d) no severe cost burden will be imposed on the trucking companies in compiling and submitting the data to TxDOT. Because no previous precedent exists for collecting statewide truck travel data through a data sharing partnership with trucking companies, it is very difficult to estimate the costs associated with this approach. It is, however, predicted that the initial costs associated with recruiting a statistically representative sample of trucking companies could be significant. The TMTA would have to be involved and compensated, and a significant public outreach effort would be required. There would also be initial costs in determining the most appropriate data collection methods for obtaining data from companies that vary not only in size but also in the level of sophistication in how the trucking companies capture this information. There will thus be upfront costs associated with developing electronic reporting options as well as less sophisticated paper-based options (i.e., questionnaires) and data entry software to facilitate the capturing of the relevant information by TxDOT or its appointed consultant. It is estimated that the costs of recruiting and developing appropriate data reporting options would be significant, but most of these costs would be incurred only initially. The recurring costs would be limited to maintaining a dialogue and relationship with the trucking companies involved and the costs associated with compiling and managing the database. This is a medium-term approach that is anticipated to be more cost-effective than the truck intercept survey approach and is worth further consideration by TxDOT. For additional detailed information about these approaches, the reader is referred to a document titled *Texas Truck Data Collection Guidebook*, which was compiled as part of this TxDOT research project.

8.6 Future Opportunities for Collecting State Truck Data

Chapter 6 highlighted a number of national initiatives for collecting freight data that might become available to state departments of transportation over the intermediate long term (i.e., five to ten years). Initiatives such as the FHWA's nationwide deployment of a universal electronic freight manifest, U.S. Customs and Border Protection's Automated Manifest System, the automated commercial environment, and the proposed International Trade Data System might result in timely and robust freight data that can be invaluable to future statewide transportation planning efforts. The chapter also mentioned the national freight data program proposed by a Transportation Research Board committee to address the general lack of freight data available to transportation planners and decision-makers. It was proposed that the database should capture the following information: (a) shipment origin and destination, (b) commodity characteristics, including weight and value, (c) mode of shipment, (d) routing and time of day, and (e) vehicle or vessel type and configuration. Although the level of state coverage is uncertain at this stage, such a national database could provide TxDOT planners with information about some variables included in the SAM. Finally, the chapter summarized the data attributes captured by current Intelligent Transportation System (ITS) technologies and highlighted a new federal ITS initiative that could potentially provide state and local transportation planners with more robust truck data over the long term. Current ITS technologies—for example inductive loop detectors, sensors, automatic vehicle classification recorders, automated vehicle identification systems, weigh-in-motion equipment, video image detection, and license plate matching systems—can collect routing, time, carrier, origin, and destination data for truck movements, but not commodity detail or truck characteristics. A new ITS initiative—dedicated short range communications (DSRC) devices—involving the FHWA, the American Association of State Highway and Transportation

Officials, a number of state departments of transportation, some of the world's largest electronic companies, toll authorities, car and truck manufacturers, and the American Trucking Association could potentially capture all the truck data required for the SAM. DSRC is a medium-range communication service intended to support both public safety (e.g., collision avoidance) and licensed private operations by means of roadside-to-vehicle and vehicle-to-vehicle communication. Devices are currently being designed that will eventually replace toll tags, CVO tags, and PrePass tags. It is predicted that these devices could be the underlying technology for electronic waybills that will make the collection of commodity data and truck characteristics feasible. At a minimum, TxDOT should stay abreast of the devices that are currently developed at the national level. Ideally, however, the agency should participate in the workshops that are hosted as part of the development of the DSRC devices to ensure that these devices allow for the collection of robust truck data required for state and local transportation planning.

8.7 Freight Forecasting Techniques

Chapter 7 provided an overview of the various freight-forecasting techniques available, ranging from simple growth factors for short-term forecasts to more complex models for long-term forecasts. In general, statewide freight demand forecasting techniques can be subdivided into three different types: causal methods, trend analysis, and qualitative methods. Causal methods include regression analysis, econometric modeling, and input-output modeling. Trend analysis or growth factors based on historical traffic trends or forecasts of economic activity are relatively simple methods to forecast freight demand over the short term. Finally, qualitative methods, such as market research methods that involve personal interviews with industry experts, panels, or focus groups, can be quite accurate in predicting, for example, major shifts in commodity flows. The truck traffic forecasting approaches employed by the FHWA for the Freight Analysis Framework and by five U.S. states were reviewed. The approaches varied significantly in terms of both the techniques and data sources used. It was found that most states (a) used the Reebie TRANSEARCH data to develop base year county commodity origin destination flow matrices, (b) developed regression equations to represent commodity productions and attractions in each county as a function of a set of explanatory socioeconomic variables, and (c) calculated future commodity productions and attractions in each county using forecasted estimates of the socioeconomic variables.

However, two states, Wisconsin and Mississippi, used trend analysis and growth factors, respectively, to develop truck traffic forecasts. The Wisconsin Freight Model developed forecasts for 2020 and five intermediate years using econometric factors derived through trend analysis. Mississippi estimated future freight flows by applying employment and population growth factors obtained from Woods and Poole Economics, Inc., to the base year commodity productions and attractions, respectively. Finally, Chapter 7 highlighted two approaches for forecasting truck data for the SAM. The first approach attempts to enhance the current regression equations used for truck forecasting in the SAM by (a) incorporating additional explanatory variables in modeling the commodity productions and attractions, (b) performing robust regression analysis for calibrating the freight generation equations, and (c) developing regression models to represent agriculture and raw material productions and attractions. The second approach forecasts truck flows using the calibrated MNL models developed as part of this research project. The relevance and robustness of the recommended approaches can, however, be determined only once the regression equations have been calibrated. Both of these approaches

will, however, take advantage of the resources already invested by TxDOT in developing a forecasting approach for the SAM and the current research project.

8.8 Concluding Remarks and Recommendations

To conclude, a number of options are available to TxDOT for collecting truck data for SAM. In the short term (i.e., the next one to three years), TxDOT can either continue to purchase the Reebe TRANSEARCH database or use the calibrated MNL models developed in this research project to estimate county-level truck data from the 2002 CFS data. The Reebe TRANSEARCH database is a relatively inexpensive source of detailed freight data at a cost of approximately \$64,000 in 2004. Questions about the robustness of the data should, however, be explored with the vendor. Alternatively, TxDOT can use the MNL models developed as part of this research project to estimate county-level truck data from the 2002 CFS database. For about half of the commodity groups, no statistically significant difference existed between the model results and the Reebe TRANSEARCH data. Because the predictive power of the MNL models are known and the inherent assumptions, limitations, and margins of error associated with the CFS data are well documented, the accuracy and reliability of the data generated from the model can be determined. Using the MNL models thus can provide TxDOT with a cost-effective alternative for obtaining truck travel data for SAM over the short term.

Over the medium term (i.e., the next three to five years), reliable truck data for Texas can be collected through an extensive program of truck intercept surveys or a data sharing initiative with trucking companies. Collecting primary data through one of these two approaches would provide TxDOT with the most robust truck travel data, but it would also be far more costly, as compared with the short-term options. Surveying ninety-six sites one day per week four times per year would cost TxDOT in excess of \$5 million using the bid prices by surveying companies for 2004 as a guideline, although some cost-saving options exist. Because there is no precedent for collecting statewide truck travel data through a data sharing partnership with trucking companies, it is more difficult to estimate the costs associated with this approach. It is, however, foreseeable that this will entail a large upfront cost arising from the recruitment of trucking companies, the development of software to minimize the burden on trucking companies in submitting the data, and finally labor costs associated with compiling and managing the database. Once these companies have entered into such a partnership, the annual costs would be limited to liaising with the companies and compiling and managing the database. Eight trucking companies that have been exposed to transportation planning have indicated their willingness to consider participating in a data-sharing program with TxDOT under certain conditions. Because it is foreseen that this approach will be more cost effective than a program of truck intercept surveys, it is recommended that TxDOT evaluates the feasibility and costs of recruiting a statistically significant sample of trucking companies through a subsequent implementation project.

In the intermediate long term (i.e., the next five to ten years), a number of national trade and ITS initiatives could potentially result in more robust freight data for transportation planning. Of these initiatives, the FHWA's development of a universal electronic freight manifest, the proposed U.S. Customs and Border Protection's trade systems, and the foreseen DSRC devices hold the most promise of providing states with more robust freight data. At a minimum, TxDOT should stay abreast of these initiatives. Ideally, however, the agency should participate in the workshops that are hosted as part of the development of, for example, the DSRC devices to

ensure that robust truck travel data are collected and made available for state transportation planning.

9. References

1. American Transportation Research Institute (2002, December). *Improving Cargo Security and Efficiency Through the Development and Testing of an Electronic Supply Chain Manifest: Phase II. Final Report*. Prepared for the Federal Aviation Administration and the Federal Highway Administration under Cooperative Agreements DTFH-61-99-X-00085 and 00-G-013.
2. Black, W. (1997). *Transport Flows in State of Indiana: Commodity Database Development and Traffic Assignment*. Phase 2, Transportation Research Center, Indiana University Bloomington.
3. Brogan, J.J., et al. (2001). *Application of a Statewide Intermodal Freight Planning Methodology*. Virginia Transportation Research Council.
4. Cambridge Systematics, Inc., Leeper, Cambridge & Campbell, Inc., Sydec, Inc., T.M. Corsi, and C.M. Grimm (1997). *A Guidebook for Forecasting Freight Transportation Demand*. NCHRP Report 388. Washington, DC: National Academy Press
5. Casavant, K.L., W.R. Gillis, D. Blankenship, and C. Howard, Jr. (1995). *Survey Methodology for Collecting Freight Truck and Destination Data*. Transportation Research Record 1477, pp. 15-25. Washington DC.
6. Cambridge Systematics, Inc. 2003. "Methods for Forecasting Statewide Freight Movements and Related Performance Measures," NCHRP Project 8-43.
7. Donnelly, R. (n.d.). *The Development of a Hybrid Microsimulation Model of Freight Flows*. Albuquerque, NM: PBConsult, Inc.
8. Federal Highway Administration (1996, September). *Quick Response Freight Manual: Final Report*. Prepared by Cambridge Systematics, Inc., Comsis Corporation, and the University of Wisconsin, Milwaukee.
9. Federal Highway Administration, USDOT, Freight Transportation Profile - Texas Freight Analysis Framework. Available online:
http://ops.fhwa.dot.gov/freight/freight_analysis/state_info/texas/tx3.pdf.
10. Federal Highway Administration (2002). *Highway Statistics 2002*. Office of Highway Policy Information. Available at <http://www.fhwa.dot.gov/policy/ohim/hs02/index.htm>.
11. Figliozi, M.A., and R. Harrison (2001, August). *Truck Trade Corridors Between the U.S. and Mexico*. Supported by Grant SWUTC/01/472840-00071-1 from the U.S. Department of Transportation, University Transportation Centers Program.

12. Gillis, W.R. (1994). *Washington State Freight Truck Origin and Destination Study: Methods, Procedures, and Data Dictionary*. Washington, DC: EWITS Research Report Number 3.
13. Gillis, W.R., E.L. Jessup, and K.L. Casavant (1995). *Movement of Freight on Washington's Highways: A Statewide Origin and Destination Study*. Washington, DC: EWITS Research Report Number 9.
14. Greene, W.H. (2000). *Econometric Analysis*. (4th Ed.). Upper Saddle River, NJ: Prentice Hall.
15. Jessup, E., K.L. Casavant, and C.T. Lawson (2004, February). *Truck Trip Data Collection Methods: Final Report*. Report Number SPR 343. Pullman, WA: Department of Agricultural and Resource Economics at the Washington State University.
16. Kockelman, K., and Y. Zhao (2001). Behavioral Distinctions: The Use of Light-Duty Trucks and Passenger Cars. *Journal of Transportation and Statistics*, 3 (3), 2001.
17. Kockelman, K. (2001). A Model for Time- and Budget-Constrained Activity Demand Analysis. *Transportation Research*, 35B (3), 255-269.
18. Lau, S.W. (1995, January). *Truck Travel Surveys: A Review of the Literature and State-of-the-Art*. Oakland, CA: Metropolitan Transportation Commission, Planning Section.
19. Lawson, C.T., A.R. Kirk, J.G. Strathman, and A. Riis (n.d.). *Making Contact: Developing an Effective Methodology to Survey the Freight Community*. Poster Presentation at the Transportation Research Board Annual Meeting, January 13-17, 2002.
20. Memmott, F.W. (1983). *Application of Statewide Freight Demand Forecasting Techniques*. NCHRP Report 260. Washington, DC: National Research Council.
21. National Research Council of the National Academies (2003). *Measuring Personal Travel and Goods Movement: A Review of the Bureau of Transportation Statistics Surveys*. Transportation Research Board Special Report 277. Washington, DC: Author.
22. Oregon Department of Transportation (1998, April). *Oregon Freight Truck Commodity Flows: Analysis and Summary*. Salem, OR: Transportation Planning Analysis Unit, Planning Section.
23. Pendyala, R.M., V.N. Shankar, and R.G. McCullough (2000). *Freight Travel Demand Modeling: Synthesis of Approaches and Development of a Framework*. Washington, DC: Transportation Research Record 1725.
24. Prozzi, J., R. Henk, J. McCray, & R. Harrison (2002, October). *Inland Ports: Planning Successful Developments*. Report Number 4083-2. Austin, TX: Center for Transportation Research at the University of Texas at Austin.

25. Strategic Freight Transportation Analysis (2002–2003). *Training Packet for Origin and Destination Survey Volunteers*. Available online:
<http://www.sfta.wsu.edu/pdf/Training.Packet.pdf>.
26. Sorratini, J.A. (2000, April). Estimating Statewide Truck Trips Using Commodity Flows and Input-Output Coefficients. *Journal of Transportation and Statistics*, 3(1), 53-67.
27. TenSiethoff, B., and K. Kockelman (2002). *Property Values and Highway Expansions: An Investigation of Timing, Size, Location, and Use Effects*. Transportation Research Record No. 1812, pp 191-200. Washington DC.
28. Texas Department of Transportation (2003, March 5). *Statewide Analysis Model 2.8.19 Freight Forecast Status Report With Technical Appendix*.
29. Texas Department of Transportation (2003). *External Station Travel Surveys* (rev. December). Specification No. TxDOT 968-85-20*, Texas Department of Transportation, General Services Division.
30. Texas Transportation Institute (2001, September). *Texas Border Crossing Travel Surveys: Executive Summary*. Prepared for the Texas Department of Transport.
31. Transportation Research Board (2003). *Special Report 276: A Concept for a National Freight Data Program*. Washington, DC: The Committee on Freight Transportation Data: A Framework for Development.
32. Triola, M.F. (2004). *Elementary Statistics* (9th ed.). Boston: Pearson/Addison Wesley.

Appendix A: Available Freight Data Sources

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
Commodity Flow Survey (CFS)	five-digit STCC (national); three-digit STCC (regional); hazardous materials also designated	State; eighty-nine NTARs based on aggregations of BEAs; foreign country for exports	Port of exit for exports	Weight and value	Mode (air/surface parcel, private and for-hire truck, rail, inland waterway, deep sea, pipeline, air and other); distances estimated; containerized shipments identified	On- and off-site facility type; equipment use by type; rail car ownership; responsibility for choice of mode (supplemental survey), publicly available
TRANSEARCH	Four-digit STCC (five-digit available for rail and water)	State; 183 U.S. BEAs (some Canadian province data); can be customized at the county or zip code level	Highway routings have been imputed from O/D data	Total weight	Mode of transport, number of transportation units	Available for purchase
Freight Transportation and Logistics Service	STCC (detail differs by mode)	Regional detail available for rail only	Not available	Cargo tonnage by mode	Equipment volumes by mode and type of equipment	Cost and rate profiles; equipment fleet size. Available for purchase
U.S. Imports / Exports of Merchandise	Ten-digit Harmonized Code, Standard Industrial Classification (SIC), BEA end-user category, and USDA agricultural product	Foreign country of origin/destination; no domestic origin/destination (beyond imputation based on district of entry/exit)	U.S. Customs district of exit (exports); or unloading and entry (imports)	Value and quantity (all modes combined); value and weight (vessel and air separately.	C.I.F. value and import freight charges for vessel and air (imports)	Number of shipment documents filed; calculated duty (imports)

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
	code.					
U.S. Exports of Domestic and Foreign Merchandise by State / Region / Port	Two-digit SIC and four-digit SITC	State or region of origin and foreign country of destination	U.S. port and district of export	Total value (all modes), total value and weight (vessel and air), containerized weight and value	Vessel, air and “all other” value; containerized weight and value (vessel and air)	
U.S. Exports by State of Origin of Movement	Two-digit SIC	State of origin and foreign country of destination	Not available	Total value (all modes), containerized and total value and weight (water and air)	Vessel, air and “all other” value; containerized weight and value (vessel & air)	
The Directory of U.S. Importers / Exporters	Actual commodity descriptions cross-referenced to ten-digit Harmonized Schedule	Address of importer/exporter (may not indicate true origin/destination), foreign country markets served/utilized	List of ports; modes utilized without allocation of volume	Total shipment value (if available)	See Routing	Bank, broker, and freight forwarder
Surface Transborder Trade-Flow Data	Two-digit TSUSA/HS (for commodity detail files only)	U.S. state/Mexican state or Canadian province of origin/destination (all files); U.S. NTAR of origin for U.S. exports (geographic detail files only)	U.S. Customs port of exit/entry for border points and Customs District for nonborder points (commodity detail files only)	Value of shipment; shipment weight (Canadian imports only); and containerized designation (U.S. imports only)	Import freight charges and containerized designation (U.S. imports only)	
U.S. Air Freight Origin Traffic Statistics (Colography)	Industry-based (four-digit SIC industry)	State, county and “market area” of origin. Destinations characterized as domestic or foreign. Colography-defined	Not available	Annual domestic and export shipment weight, value and number of shipments (with weight and number	Shipment size categories corresponding to standard market classifications in air freight	Total employment and number of plants (total and by employment size) by area

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
		“market areas” are aggregates of counties corresponding to local hinterland of U.S. airports or customized set of counties.		of shipments shown by shipment size categories)	industry (express, heavy freight)	
Airport Activity Statistics of Certified Route Air Carriers	Not available	Airport of enplanement	Carrier with no routing information	Enplaned tons of freight (express and nonexpress) and mail (priority, nonpriority and foreign)	Departures by service type (scheduled or nonscheduled) and equipment and type	
Worldwide (North American) Airport Traffic Report	Freight, express or mail	Domestic/international flight	Airport	Type of shipment (freight and express or mail) plus total international freight plus mail	Total weight by airport and direction (enplaned or deplaned); airport domestic and international aircraft operations by aircraft and type	
ICC Carload Waybill Sample	Seven-digit STCC on MF; two- to five-digit STCC on PUF excluding hazardous materials (STCC 49) and bulk materials in boxcars (STCC 50) which are classified separately.	O/D of rail movement identified by BEA Region. Intermodal, import, export and mini-bridge shipments are flagged. MF also contains six-digit Standard Point Location Code (SPLC) and a Freight Station Accounting Code	Interchange states and number of interchanges (PUF); Full railroad and station itinerary (MF)	Billed and actual tons, carloads, trailers, containers, and revenue (sample & expanded universe totals); date of shipment/waybill	Equipment type, shipment and expanded revenue by type (freight, transit, miscellaneous), short line miles, number of interchanges, number of intermodal units; carrier &	

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
					equipment type, design, capacity, dimensions. Short line miles is the shortest rail distance.	
Freight Commodity Statistics	Two-, three-, four-, and five-digit STCC; some shipments cannot be classified at the five-digit level based on available documentation or mixed loadings, so higher level groupings may not be fully described at disaggregated levels	Not available	Not available	Total tons for the following type of shipment: 1. Originated and terminated 2. Originated and delivered to another carrier 3. Received and terminated 4. Received and delivered to another carrier	Freight revenue and carloads by commodity for originated freight, terminal freight and total freight	None
North American Trucking Survey (NATS)	STCC (three-digit for major commodities)	City and state of O/D	Not available	Weight in tons	Trailer type	Annual VMT of driver; operator characteristics (private, for-hire, owner-operator).
LTL Commodity and Market Flow Data Base	Categorized by service type (standard, nonstandard delivery time, special equipment/handling)	O/D zip codes or foreign area (Canada, Mexico, Asia, Europe or Other)	Mileage	Weight, number of shipments, and number of pieces	Ton-miles, revenue, service type, intermodal and interline indication	Intermodal shipments are identified

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
Truck Inventory and Use Survey (TIUS)	Percentage of annual mileage for twenty-six commodity categories, including nonfreight activity (personal, idle, and empty haul use) and (separately) for seventeen categories of hazardous materials.	Not available	Not available	Not available	Percentage of annual miles outside of designated "base" state and percent by range of operation categories	Type of business in which vehicle was used. Vehicle (make, year, dimensions, body/trailer type, capacity, axle and operating configuration, equipment, maintenance) acquisition/disposition (year, method, lease/ownership) utilization (annual and lifetime mileage, fuel efficiency, state of operation, type of use, Hazmat activity, commodity types, accident incidence)
Nationwide Truck Activity and Commodity Survey (NTACS)	Twenty-six TIUS commodity categories (including empty) plus Hazmat categories	Sample day cargo load and discharge patterns* -stop location -type of place (e.g., warehouse, port)	Sample day cargo routing patterns* -detailed stop locations -type of stop activity (e.g., pick up, delivery) -arrival and departure time	1987 percentage of total mileage by commodity (TIUS) 1990 sample day weight by commodity and load/discharge stop*	Annual -weeks of operation -annual mileage -number of states -top three states of operation* Canada/Mexico* -type of use -configuration and body/trailer	Vehicle (model year, ownership, operation) *Master File Only

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
					type Sample Day -ton-miles for top commodity -days of week and time of day use -number of stops -type of use -mileage -type of highways -fuel use and tolls -dimensions, capacity and load	

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
Port Import / Export Reporting Service (PIERS)	Six-digit Harmonized, seven-digit PIERS Commodity code (loosely based on 1979 TSUSA), and actual manifest/bill of lading description	U.S. shipper/consignee and foreign shipper (import only) -name -U.S. city of O/D -foreign country of import/export (city for export)	U.S. port of loading/unloading -Foreign port of ultimate O/D -Foreign port of transshipment -No inland mode designation	Shipment weight and value -Package type and quantity	Carrier and vessel name -Container size, number and estimate of cubic volume utilized -Package type	U.S. port date, linkage to other company information for importers/exporters is available
Waterborne Commerce and Vessel Statistics	Four-digit Commodity Classification for Domestic Waterborne Commerce	Dock-to-dock flows (master tape); port/harbor/channel segment throughput (port summary); state-to-state (public domain database). O/D of vessel may be inferred from flow type categories (e.g., internal, coastwise). No foreign country detail for international shipments	Data is provided for specific route elements; specific routing patterns are not available beyond inferred routing based on O/D combination	Shipping weight (tons)	Number of vessels by direction, type and draft	
World Sea Trade Service	Twenty SITC-based commodity groups (additional detail available for certain countries)	Foreign country of O/D (based on ports of lading and discharge; transshipment activity not identified)	Trade routes as defined by coast/country/region pairs at various levels of detail; port detail is available for certain trade routes	Total weight and container loads	Number of container loads	
Exports from Manufacturing Establishments	Three-digit SIC industry group (detail may be suppressed for confidentiality)	State of production	Not available	Shipment value (F.O.B.) for direct and supporting exports	Not available	Employment for direct and supporting exports

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
Fresh Fruit and Vegetable Shipments by Commodities, States and Months	Individual fruits and vegetables with some grouping of minor commodities and mixed load shipments; domestic and export commodities are listed separately	U.S. state (with four-district detail for California) or foreign country of origin; domestic or export destination group	Not available	Cargo weight by month and year	Mode of transport for domestic shipments (rail refrigerated cars, piggyback, truck, air and water)	
Fresh Fruit and Vegetable Arrival Totals for Twenty-Three Cities	Individual fruits and vegetables with some grouping of minor commodities and mixed load shipments	U.S. state or foreign country origin; city of destination	Not available	Cargo weight by month and year	Mode of transport (rail, truck, air, and water)	
Quarterly Coal Report	Coal (with some data broken in categories defined by origin or physical properties such as BTU content)	U.S. State or foreign country of origin and (separately) destination country or U.S./Canadian sector (electric generation, coke plants, industrial plants, residential and commercial); no O/D pairs	Customs District for imports and exports	Weight	Principal mode (rail, inland waterway, Great Lakes, ocean port, truck, slurry); no intermodal designations	
Petroleum Supply Monthly	Separate statistics for crude oil and primary petroleum products	Foreign country for imports and exports; U.S. O/D inferable based on port routing	Imports: traffic by country, multi-state Petroleum Administration for Defense (PAD) of entry and commodity. Exports: traffic by commodity and country or PAD of exit. Domestic	Volume in barrels	Mode (Pipeline, tanker and barge)	

Data Source	Commodity	Origin/Destination	Routing	Shipment	Transport	Other
			PAD-to-PAD traffic flows by mode (may include foreign transshipments)			

Source: NCHRP Report 388, 1997

Appendix B: Mississippi Intermodal Freight Model

The Mississippi Department of Transportation (MDOT), in conjunction with the Mississippi State University, developed a methodology to estimate county-level data from the 1997 CFS data. The method used to estimate county-to-state, state-to-county, and county-to-county flows are highlighted in this section of the report.

County-to-State Flows

Freight flow data from Mississippi to other U.S states captured in the CFS database and county employment data captured by the Census Bureau's County Business Patterns data were used to generate county-to-state flows as follows:

$$TO_{ij}^k = \frac{E_i^k}{E_M^k} * TO_{Mj}^k$$

where

TO_{ij}^k = Total flows of commodity k from county i to attraction state j ,

E_i^k = Employment in county i in production occupations of commodity k ,

E_M^k = Employment in Mississippi in production occupations of commodity k , and

TO_{Mj}^k = Total flows of commodity k from Mississippi to attraction state j .

State-to-County Flows

Freight flows from U.S. states to Mississippi captured in the CFS database and the U.S. Bureau of the Census county population data were used to generate state-to-county flows as follows:

$$TD_{ji}^k = \frac{PN_j^k}{PN_M^k} * TD_{Mi}^k$$

where

TD_{ji}^k = Total attractions of commodity k to county j from production state i ,

PN_j^k = Population of county j ,

PN_M^k = Population of Mississippi, and

TD_{Mi}^k = Total attractions of commodity k to Mississippi from production state i .

County-to-County Flows

County productions destined for Mississippi are computed from the CFS Mississippi-to-Mississippi flows on the basis of the fractional county employment as follows:

$$TO_{iM}^k = \frac{E_i^k}{E_M^k} * TO_{MM}^k$$

where

TO_{iM}^k = Total productions of commodity k in county i destined for Mississippi,

E_i^k = Employment in county i in production occupations of commodity k ,

E_M^k = Employment in Mississippi in production occupations of commodity k , and

TO_{MM}^k = Total productions of commodity k in Mississippi destined for Mississippi.

County attractions originating within Mississippi are computed from the CFS Mississippi-to-Mississippi flows on the basis of the fractional county population as follows:

$$TD_{jM}^k = \frac{PN_j}{PN_M} * TD_{MM}^k$$

where

TD_{jM}^k = Total attractions in county j of commodity k produced in Mississippi,

PN_j = Population of county j ,

PN_M = Population of Mississippi, and

TD_{MM}^k = Total attractions in Mississippi of commodity k produced in Mississippi.

Once the county productions and attractions of commodities were determined, the gravity model was used to generate county-to-county freight flows. The inverse square of the intercounty distance (d_{ij}^{-2}) was specified as the impedance factor for the gravity model (NCHRP, 1998). Freight flows of commodity k from county i to county j was thus determined as follows:

$$T_{ij}^k = TO_{iM}^k * \frac{TD_{jM}^k d_{ij}^{-2}}{\sum_{j=1}^n TD_{jM}^k d_{ij}^{-2}}$$

where

T_{ij}^k = Freight flows from county i to county j for commodity k ,

TO_{iM}^k = Total productions of commodity k in county i destined for Mississippi,

TD_{jM}^k = Total attractions in county j of commodity k produced in Mississippi, and

d_{ij}^{-2} = Impedance factor for gravity model.

Appendix C: List of Acronyms

ACE	Automated Commercial Environment
AMS	Automated Manifest System
ANSI	American National Standards Institute
ATSSA	American Traffic Safety Services Association
AVC	Automatic Vehicle Classification
AVI	Automated Vehicle Identification
AVL	Automatic Vehicle Location
BEA	Bureau of Economic Analysis
BEAs	Business Economic Areas
CBP	County Business Patterns
CEDDS	Complete Economic and Demographic Data Source
CFS	Commodity Flow Survey
CVO	Commercial Vehicle Operations
DSRC	Dedicated Short Range Communications
EC	Electronic Credential
ESCM	Electronic Supply Chain Manifest
EWITS	Eastern Washington Intermodal Transportation Study
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
FISHFM	Florida Intermodal Statewide Highway Freight Model
GPS	Global Positioning Systems
HAZMAT	Hazardous Materials
HS	Harmonized Schedule of Foreign Trade

ICC	Interstate Commerce Commission
IFS	North Central Texas Council of Government's Intermodal Freight and Safety Committee
IMPLAN	IMpact analysis for PLANning
I-O	Input-Output
ISTEA	Intermodal Surface Transportation Efficiency Act
ITDS	International Trade Data System
ITS	Intelligent Transportation System
LATTS	Latin American Trade and Transportation Study
LPM	License Plate Matching
MDOT	Mississippi Department of Transportation
MIG	Minnesota IMPLAN Group
MNL	Multinomial Logit Model
NAFTA	North American Free Trade Agreement
NAICS	North American Industry Classification System
NATS	North American Trucking Survey
NTACS	National Truck Activity and Commodity Survey
NTARs	National Transportation Analysis Regions
NYSDOT	New York State Department of Transportation
O/Ds	Origins and Destinations
OES	Occupational Employment Statistics
OLS	Ordinary Least Squares
PAD	Petroleum Administration for Defense
PIERS	Port Import/Export Reporting Service
REMI	Regional Economic Models Inc.

SAM	Statewide Analysis Model
SFTA	Strategic Freight Transportation Analysis
SIC	Standard Industrial Classification
SITC	Standard International Trade Classification
SPLC	Standard Point Location Code
SSL	Secure Socket Layers
STCC	Standard Transportation Commodity Codes
TAZs	Traffic Analysis Zones
TEA-21	Transportation Equity Act for the Twenty-First Century
TEEX	Texas Engineering Extension Service
TIUS	Truck Inventory and Use Survey
TMTA	Texas Motor Transportation Association
TxDOT	Texas Department of Transportation
VIUS	Vehicle Inventory and Use Survey
VTRC	Virginia Transportation Research Council
WEFA	Wharton Economic Forecasting Associates
WIM	Weigh-In-Motion