Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-01/0-1833-4		

4. Title and Subtitle	5. Report Date
FREIGHT MODAL SPLIT MODELING: CONCEPTUAL FRAMEWORK, MODEL STRUCTURE, AND DATA SOURCES	August 2000
7. Authors Aruna Sivakumar and Chandra R. Bhat	6. Performing Organization Code 8. Performing Organization Report No. 0-1833-4
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. Project 0-1833
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Transfer Section/Construction Division P.O. Box 5080 Austin, TX 78763-5080	13. Type of Report and Period covered Research Report (01/2000 to 08/2000) 14. Sponsoring Agency Code

15. Supplementary Notes

Project conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration, and the Texas Department of Transportation.

16. Abstract

This report discusses the modal split model for freight movements within the context of a larger model system that can forecast the effects of port expansions, market changes, and network changes on the statewide transportation network. Specifically, the report focuses on the conceptual framework for the modal split model, the model structure, and the data sources and assembly procedures for model estimation.

17. Key Words:	18. Distribution Statement
Modal split, mode choice, freight flow, Reebie data, quasi-likelihood estimation, road and rail, truck and rail, fractional split, logit, freight interchanges, county-to-county, STCC2	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	26	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FREIGHT MODAL SPLIT MODELING:
CONCEPTUAL FRAMEWORK, MODEL STRUCTURE, AND DATA SOURCES

by

Aruna Sivakumar and Chandra R. Bhat

Research Report Number 0-1833-4

Research Project 0-1833
Infrastructure Impacts and Operational Requirements Associated with the Next Generation Container Ships (Megaships) on the Texas Transportation System

Conducted for the

TEXAS DEPARTMENT OF TRANSPORTATION

in cooperation with the

U.S. DEPARTMENT OF TRANSPORTATION

Federal Highway Administration

by the

CENTER FOR TRANSPORTATION RESEARCH

Bureau of Engineering Research

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2000

DISCLAIMERS

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 views or policies of either the Federal Highway Administration or the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

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 and useful improvement thereof, or any variety of plant, which is or may be patent able under the patent laws of the United States of America or any foreign country.

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES

Chandra R. Bhat
Research Supervisor

ACKNOWLEDGMENTS

The research supervisor wishes to recognize the support and participation of C. Michael Walton Professor (CTR), Robert Harrison, Senior Research Scientist (CTR); Miguel Figliozzi, graduate student in transportation; and Michael Bomba, LBJ School graduate student intern. Also significant to the accomplishment of this product has been the involvement and direction of the Texas Department of Transportation (TxDOT) project steering committee, which includes project director Raul Cantu (TPP) and project monitoring committee members Jim Randall (MMO), Kelly Kirkland (MMO), Chris Olavson (HOU), and Carol Nixon (HOU).

Chandra R. Bhat, Research Supervisor

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The introduction of megaship technology into maritime container freight trade to meet the growing demand for international economic expansion is likely to influence the landside traffic patterns in and around Texas ports. Proper consideration of, and planning for, these landside traffic impacts is important to ensure an efficient transportation system that is able to accommodate the changing traffic patterns. In this context, it becomes important to develop a modeling system that can forecast the effects of port expansions, market changes, and network changes on the statewide transportation network. Figure 1 shows the framework for such a modeling system. The system takes the following as inputs: a) the commodity types and trade volumes (throughputs) moving into the port under consideration, b) the markets served by the port and the socio-demographics of the markets, and c) the level-of-service offered by alternative modes from the port to the markets served by the port. The desired outputs from the system are the freight-related traffic changes on the transportation network resulting from changes in one or more inputs.

The modeling system in Figure 1 comprises two main components. The first is a modal split model that predicts the fraction of freight moved by each of the major transportation modes between the port under consideration and each market served by the port. The second is a network assignment model that translates the highway freight tonnage into equivalent highway traffic and assigns this traffic to the state highway transportation network.

The focus of this technical report is on the modal split component of the modeling system. Specifically, the report focuses on the conceptual framework for the modal split model (section 2), the structure for the model (section 3), and the data sources and data assembly procedures for model estimation (section 4). The final section concludes the report.

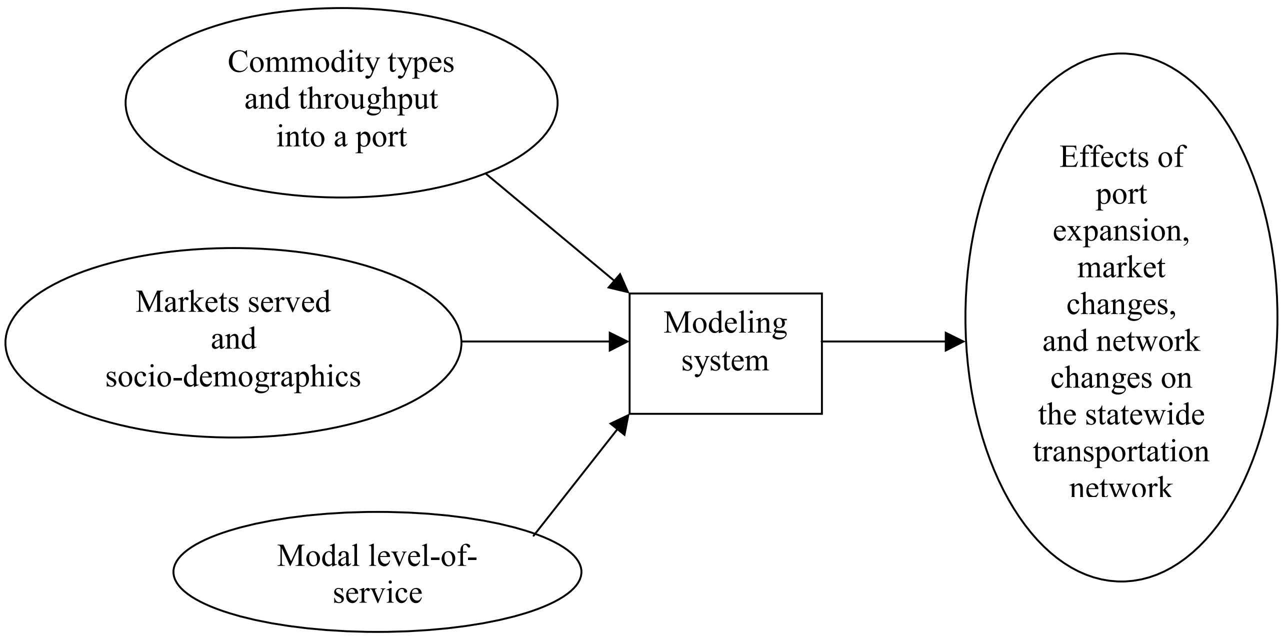

Fig. 1. Framework for modeling system

2. THE CONCEPTUAL FRAMEWORK

The objective of the modal split model, as discussed in the previous section, is to predict the fraction of freight moved by each of the transportation modes between a port and the markets served by the port given the following: a) projected trade volume of the port, b) the projected markets, and the socio-demographics of the markets, served by the port, and c) the transportation level-of-service offered by competing travel modes. There are two steps involved in this prediction process. The first step is to *estimate* the influence of trade volumes, market socio-demographics, and level-of-service of alternative modes on modal split. The second step is to *apply* the estimated model to obtain the freight volumes transported by each mode in response to any given combination of the input variables.

2.1. *Estimation of Mode Split Model*

The estimation of the modal split model requires data on current modal splits and the corresponding values of the input variables. However, such data is not available from Texas ports. Consequently, the research team used the Reebie freight database to estimate the effects of input variables on modal split. The Reebie data provides commodity-type-specific information on freight modal splits for each county-to-county pair within Texas and between Texas and several spatially aggregate external stations outside Texas. It also provides information on total freight tonnage movements. Our estimation procedure will extract the relevant freight movement data from the Reebie database that is specific to the commodity types flowing in and out of Texas ports. This information will be supplemented with county and external station socio-demographic data (i.e., population, employment level, average income, and other economic indicators) and with network level-of-service attributes to estimate a modal split model (see Fig. 2). The final result from this step is a commodity-type-specific modal split model with estimated sensitivities to total tonnage of freight being moved, county/external station socio-demographics, and the network level-of-service between counties by alternative modes.

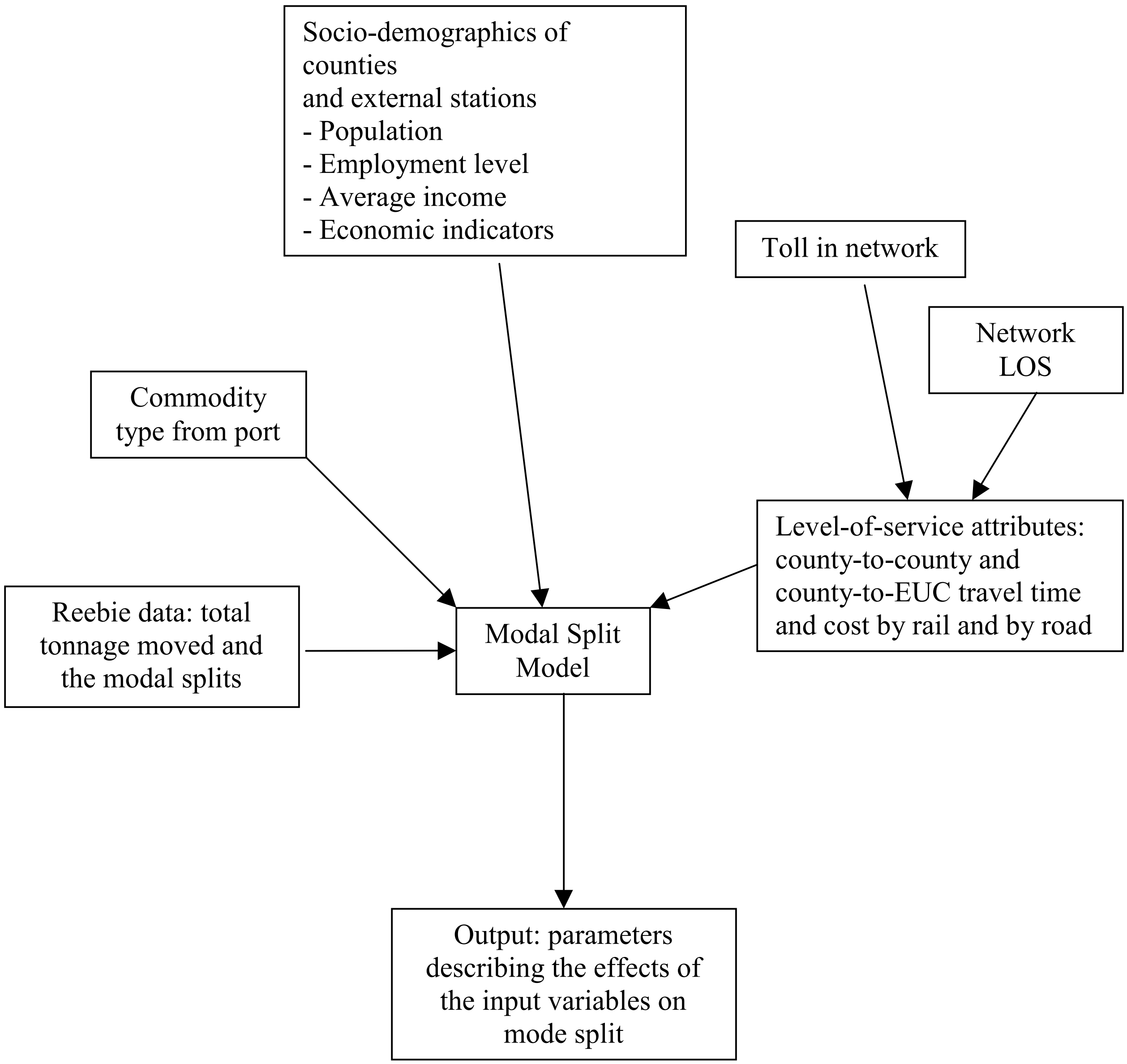

Fig. 2. Estimation of modal split model

The approach outlined above can be used to estimate the effects of input variables on splits among any number of travel modes. However, the two modes that dominate landside freight movement are the highway (truck) mode and the rail mode. Hence, the estimation of the mode choice model is restricted to these two modes in the current project.

2.2. *Application of the Mode Split Model*

The application of the estimated mode split model to any Texas port is conceptually straightforward, as shown in Figure 3. The inputs required at this stage are: a) trade volumes or "throughput" of the port (both outbound and inbound) by commodity type, b) the markets served by the port and their respective socio-demographics, and c) the level-of-service attributes for travel by truck and rail between the port and each market. These inputs can correspond to a projected situation at a future time. The output will be the predicted mode splits from (to) the port to (from) the destination (origin) counties/external stations corresponding to the projected situation implied by the input values. The modal splits can then be further translated into the predicted rail and truck freight movements between the port and each market by commodity type.

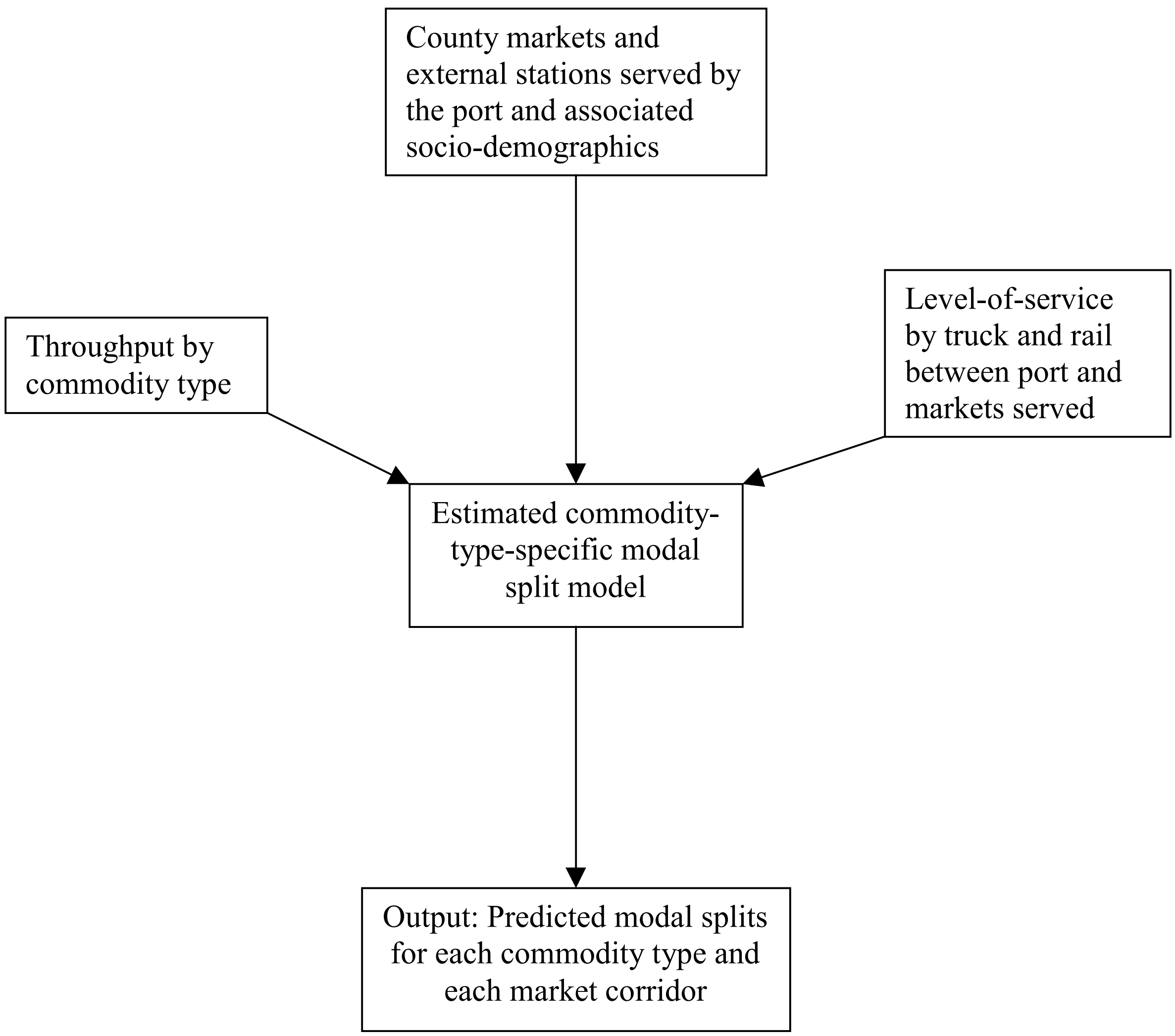

Fig. 3. Application of the modal split model

3. MODEL FORMULATION — FRACTIONAL SPLIT MODEL STRUCTURE

3.1. *General Background*

The dependent variable in the freight modal split analysis framework discussed in section 2.1 is the fraction of total county-to-county freight volume transported by the rail mode (the reader will note that the sum of the rail and highway fractions should equal one for each county-to-county pair and so it suffices to focus on one of the two fractions). By construction, the rail fraction is bounded by the values of zero and one. Further, the rail fraction may take the boundary values of zero or one; i.e., all the freight for some county-to-county pairs may be transported by road and all the freight for some other county-to-county pairs may be transported by rail.

Mathematically, let y_q be the rail fraction of freight tonnage moved between the qth county-to-county pair (we will develop commodity-type-specific modal split models and so do not use an index for commodity type in the following presentation). Let this fraction be a function of a vector x_q of relevant explanatory variables, including the network level-of-service variables for each of the modes, and the characteristics associated with the qth origin-destination pair. A common approach to analyzing fractional dependent variables is to model the log-odds ratio as a linear function (for example, see Bhat and Misra [1]):

$$E\left[\log \frac{y_q}{1-y_q}\right] = \beta' x_q \tag{1}$$

where β is a parameter vector representing the effect of exogenous variables on rail mode share.

The specification in equation (1) is attractive since the transformed dependent variable in the regression is unbounded and can take values anywhere on the real line as y_q varies between 0 and 1. Thus, a linear regression is appropriate. However, as pointed out by Papke and Wooldridge [2], the specification has at least two major problems. First, the dependent variable is undefined when the fraction of freight moved by rail is zero or one. If the number of county-to-county pairs for which the boundary conditions prevail is few, small arbitrary adjustments may be made before computing the log-odds ratio without significantly affecting the estimated parameters. However, if there are several county-to-county pairs for which the boundary conditions prevail, the adjustments can have a substantial impact on estimation. In our analysis, the fraction of freight moved by rail is zero for a large percentage of county-to-county pairs. Hence, the specification in equation (1) is inadequate. A second problem with the specification in equation (1) is that, even if the econometric specification in equation (1) is appropriate and well defined, one cannot obtain $E\left(y_q \mid x_q\right)$ without additional assumptions about the distribution of the residuals, $u_q = \log\left[y_q/(1-y_q)\right] - \beta' x_q$. If a distribution is assumed or estimated, then $E\left(y_q \mid x_q\right)$ may be computed by first obtaining the conditional (on residuals) expected value for each fraction and then unconditioning out the residuals by integrating over the distribution of the assumed or estimated distribution for the residuals (see Bhat [3] for an application of this method). However, this approach is either nonrobust (if an incorrect parametric distribution is assumed) or cumbersome (if a nonparametric distribution for the residuals is estimated).

3.2. *Quasi-Likelihood Estimation*

In the current modal split analysis, we use the binary fractional split model proposed by Papke and Wooldridge [2]. The approach does not need any ad hoc adjustment for boundary values of the dependent variable fractions and directly specifies a model for $E\left(y_q \mid x_q\right)$. At the same time, the approach is easy to implement and is robust because we make no assumptions about the distribution of y_q conditional on x_q. The focus is on consistent estimation of the parameters appearing in the conditional mean specification $E\left(y_q \mid x_q\right)$ and on consistent, asymptotically robust, estimation of the standard errors of the conditional mean parameters.

Consider the following econometric specification:

$$E\left(y_q \mid x_q\right) = G\left(\beta, x_q\right),\ 0 < G(.) < 1 \tag{2}$$

where G(.) is a predetermined function and the properties specified above for it ensure that the predicted rail freight fraction will lie in the interval (0,1). The econometric model in equation (2) is well defined even if y_q takes the value of 0 or 1 with positive probability.

The β parameter vector in the conditional mean model of equation (2) is estimated by maximizing a likelihood function associated with a family of probability distributions that does not necessarily contain the true underlying conditional distribution of y_q given the x_q values. The label "quasi-likelihood estimation" is used for such estimations (see Gourieroux et al. [4]). Specifically, we use the following log-likelihood function in the quasi-estimation:

$$L_q(\beta) = y_q \log[G(\beta, x_q)] + (1 - y_q)\log[1 - G(\beta, x_q)]. \tag{3}$$

We base our inference only on the conditional mean specification of equation (2) and propose consistent and asymptotically robust inference for the conditional mean parameter vector β. As indicated by Papke and Wooldridge [2], this inference can be achieved by computing the asymptotic variance-covariance matrix of β as $H^{-1}\Delta H^{-1}$, where H is the Hessian and Δ is the cross-product matrix of the gradients (H and Δ are evaluated at the estimated parameter values).

A final model structure issue concerns the specification of the functional form for G in the conditional mean specification of equation (2). We use a logit functional form for G because this structure is easy to program and implement. In this structure, we write:

$$G(\beta, x_q) = \frac{1}{1 + e^{-\beta' x_q}} \tag{4}$$

4. DATA PREPARATION

4.1. *Data Sources*

Several data sources will be used in the estimation phase of the modal split model. These include: a) the Reebie TRANSEARCH Freight Database, b) TRANSCAD geographic maps and datasets, c) County Business Patterns compiled by the U.S.

Census Bureau, d) the U.S. Census Bureau Population projections, and e) the U.S. Bureau of Economic Analysis. Each of these data sources is briefly discussed here.

The Reebie TRANSEARCH Freight Database of 1996 is a proprietary database of county-to-county freight movements throughout the United States. It is compiled and produced annually, based on surveys of major shippers and carriers in the freight industry. The database comprises four files, each providing intercounty freight movements by one of four major transportation modes (road, rail, water, and air). Within each of the four files, the intercounty freight movements are disaggregated into fifty commodity types based on the STCC2 classification code. (STCC2 is the two-digit Standard Transportation Commodity Classification developed and maintained by the Association of American Railroads). The part of the Reebie data relevant to Texas freight movements includes county-to-county flow within Texas and flows between Texas counties and each of several aggregate zones external to Texas (these aggregate zones are referred to as external unit codes [EUC] and are listed in Appendix I). This Texas-related freight flow data (hereafter referred to as the Texas flow data) forms the basis for computing the fractional mode split by rail for each county-to-county movement (and for each county-EUC interchange). This rail modal split constitutes the dependent variable in the current analysis. The data also provides information on total freight volumes for each interchange, which is an explanatory variable in the analysis.

The TRANSCAD geographic maps and datasets are used to determine the centroidal distances between Texas counties and the distances between Texas counties and the external unit codes. The geographic areas of the counties and EUCs are also computed from these maps.

The County Business Pattern database is maintained by the U.S. Census Bureau and has been published every year since 1946 at the county, state, and national levels. The data provides county-specific information on economic indicators, including establishment counts by institution size and mid-March employment figures. This dataset represents our primary source of socio-economic information on Texas counties. Details of the sampling procedure and methods used to compile this information are available on the Web site of the U.S. Census Bureau (http://www.census.gov/pub/epcd/cbp/download/cbpdownload.html).

The U.S. Census Bureau population projections provide population predictions based on decennial census counts. These predictions use the population counts obtained from the most recent census survey as the base and can update these base counts to any future year based on assumptions about future births, deaths, international migration, and domestic migration. The population figures used in our analysis correspond to 1996, since this was the year the Reebie data was compiled.

The Bureau of Economic Analysis (BEA) is an agency of the U.S. Department of Commerce. The mission of BEA is "to produce and disseminate accurate, timely, relevant, and cost-effective economic accounts statistics at the regional, national and international level" (source: Bureau of Economic Analysis Web site http://www.bea.doc.gov). The regional economic accounts provide estimates and analyses of personal income, population, and employment for regions, states, metropolitan areas, and counties. The county level estimates of personal income and employee count for 1996 are used in the current analysis.

4.2. *Data Assembly*

The objective of data assembly is to assemble the dependent variable and the independent variables for analysis in a single data file from the various data sources discussed in the previous section. Several steps were involved in this process, as we discuss next.

The first step collapsed the fifty commodity types in the Texas flow data into the following seven aggregate commodity types: a) agricultural and related products; b) hazardous materials; c) construction materials; d) food and related products; e) manufacturing products; f) machinery and equipment, and g) mixed freight shipments.

The second step screened the Texas flow data to ensure that the origin and/or destination end of each freight interchange was a Texas county. The records of several freight interchanges did not satisfy this condition and were therefore removed from the data.

The third step appended the Texas county-to-county and Texas county-to-EUC distances (generated from the TRANSCAD maps) to the Texas flow database based on the origin and destination identifiers of each freight interchange. Intrazonal freight interchanges were assigned a distance equal to half the distance from that zone to its closest neighboring zone.

The fourth step similarly mapped the relevant county-specific or EUC-specific socio-demographic and economic indicators (obtained from the supplementary data sources discussed in the previous section) to each freight interchange based on the origin and destination of the interchange.

The final sample comprises 108,458 freight interchanges, of which 72,986 cases (67.3%) represent Texas county-to-county flows, 17,715 cases (16.3%) represent Texas county-to-EUC flows, and 17,757 cases (16.4%) represent EUC-to-Texas county flows. The sample includes the following variables for each interchange: origin zone; destination zone; aggregate commodity code; freight tonnage by road and rail; total shipment size for the given zonal pair; fraction of freight by road and rail; distance between the zones; intrazonal dummy variable; and population, area, population density, personal income, and economic indicators of the origin and destination zones.

4.3. *Sample Description*

The descriptive statistics of the rail mode fraction (the independent variable in the analysis) and of the shipment size and distance of haul (two independent variables in the analysis) are provided in Table 1. As expected, on average, the rail mode fraction is much lower than the highway mode fraction. The rail fractional split appears to be highest for agricultural and related products. The percentage of observations for which the fractional split is very close to the boundary value of 0 is very high and ranges from 85.1% for the agricultural and related products category to 99.5% for the machinery and equipment commodity category. Thus, the log-odds ratio approach, which cannot handle the boundary value of 0 for the dependent variable, would seem inappropriate in the current empirical analysis. The specification in equation (2) can accommodate such boundary values and is the most appropriate technique.

An examination of the rail fraction with the distance of haul and shipment size reveals some interesting patterns. Specifically, the rail fraction is highest for the agricultural and related products category; this category also has the largest distance of haul. This pattern suggests a possible positive effect of haul distance on rail fraction. Another finding is that the rail fraction is lowest for the machinery and equipment category; this category also has the smallest shipment size.

Table 1: Average rail fraction, distance of haul, and shipment size by commodity type

Commodity Type	*Sample Size*	*Rail Fraction*		*Distance of Haul*		*Shipment Size*	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
#1. Agricultural and Related Products	2,961	.143	.347	302.876	156.783	2163.762	15719.645
#2. Hazardous Materials	18,120	.048	.192	265.608	147.463	8840.602	169510.44
#3. Construction Materials	19,178	.032	.168	240.427	136.818	7516.231	86994.864
#4. Food and Related Products	14,706	.023	.136	262.026	145.592	6430.080	66179.325
#5. Manufacturing Products	23,227	.021	.131	241.148	133.386	1204.192	19520.104
#6. Machinery and Equipment	15,161	.003	.044	247.838	140.347	206.492	3346.086
#7. Mixed Freight Shipments	15,105	.010	.091	269.603	149.124	6494.910	93861.285

5. CONCLUSION

This report discusses the modal split model for freight movements within the context of a larger model system that can forecast the effects of port expansions, market changes, and network changes on the statewide transportation network. Specifically, the report focuses on the conceptual framework for the modal split model, the model structure, and the data sources and assembly procedures for model estimation. Model estimation and specification testing are ongoing efforts that will be presented in a subsequent report.

APPENDIX I

Table 2: External unit code (EUC) description

External Unit Code	Literal Description	Description (as on map)
05000	AR (Net of MEM)	Arkansas
22000	LA (Net of NOL)	Louisiana state (excluding New Orleans)
35000	NM	New Mexico
40000	OK	Oklahoma
48000	TX	Texas
BE064	BEA Chicago, IL	Chicago
BE073	BEA Memphis, TN	Memphis
BE083	BEA New Orleans, LA	New Orleans
BE096	BEA St. Louis, MO	St. Louis
BE099	BEA Kansas City, MO	Kansas City
BE141	BEA Denver, CO	Denver
BE160	BEA Los Angeles, CA	Los Angeles
R1NEN	Region 1: New England	New England
R2MAT	Region 2: Mid Atlantic	Mid Atlantic
R3ENC	Region 3: East North Central (Net)	East North Central
R4WNC	Region 4: West North Central (Net)	West North Central
R5SAT	Region 5: South Atlantic	South Atlantic
R6ESC	Region 6: East South Central (Net)	East South Central
R8MTN	Region 8: Mountain (Net)	Mountain
R9PAC	Region 9: Pacific (Net)	Pacific (excluding Los Angeles)

REFERENCES

1. Bhat, C. R., and R. Misra. "Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends," *Transportation*, Vol. 26, No. 2, 1999, 193–209.
2. Papke, L. E., and J. M. Wooldridge. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates," *Journal of Applied Econometrics*, 11, 1996, pp. 619–632.
3. Bhat, C. R. *Modeling the Commute Activity-Travel Pattern of Workers: Formulation and Empirical Analysis*, Technical Paper, Department of Civil Engineering, The University of Texas at Austin.
4. Gourieroux, C., A. Monfort, and A. Trognon. "Pseudo maximum likelihood methods: theory," *Econometrica*, 52, 1984, pp. 681–700.