An Analysis of Origin-Destination Freight Flows Using a Structural Equations Modeling Framework

Srikanth Vaddepalli
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of South Florida, ENB118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: (813) 974-8726; Fax: (813) 974-2957
Email: svaddepa@eng.usf.edu

Siva S. Jonnavithula
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of South Florida, ENB118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: (813) 974-8726; Fax: (813) 974-2957
Email: sjonnavi@eng.usf.edu

Ram M. Pendyala, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of South Florida, ENB118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: (813) 974-1084; Fax: (813) 974-2957
Email: pendyala@eng.usf.edu

Robert G. McCullough, P.E.
Systems Planning Office
Florida Department of Transportation
605 Suwannee Street, MS19
Tallahassee, FL 32399-0450
Ph: (850) 414-4931; Fax: (850) 921-6361
Email: bob.mccullough@dot.state.fl.us

Accepted for Presentation and Resubmitted for Publication
83rd Annual Meeting of the Transportation Research Board
January 11-15, 2004, Washington, D.C.

Revised November 2003

TRB Paper No.: 04-4625
Word Count: 6736 text + 250 x 6 tables/figures = 8236

Abstract

The modeling of freight travel demand has gained increasing attention in the recent past due to the importance of efficient and safe freight transportation to regional economic growth. Despite the attention paid to the modeling of freight travel demand, advances in modeling methods and the development of practical tools for forecasting freight flows have been limited. This paper attempts to make a contribution in this context by developing a structural equations modeling framework that can be applied to the modeling of freight travel demand using data contained in readily available commercial databases such as the Reebie Transearch database and the InfoUSA employer database. The model system is estimated at the level of the individual zip code and can be used to estimate freight flows by commodity by major modes between origin-destination zip code pairs. The paper describes the database development, model framework, sample model estimation results, and directions for further model development in the future.

Keywords: Freight transportation, population, employment, structural equations, freight data

1. INTRODUCTION

Freight transportation is a critical element in the overall demand-supply chain of commodities and services in a region. Freight transportation is critically tied to the economic growth and well-being of a region that freight transportation planning has become a major focus of transportation planning in many areas around the country and world. In the United States, federal legislation such as the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) and the Transportation Equity Act for the 21st Century (TEA-21), and a host of state-level initiatives have provided the impetus for undertaking comprehensive freight transportation planning and mobility strategies.

An important part of freight transportation planning is the ability to quantify and predict freight flows between origin-destination pairs in a region of interest. These origin-destination pairs may be traffic analysis zones, census tracts, zip codes, cities, counties, or even states depending on the particular freight transportation planning context of interest. As with any transportation modeling effort, the ability to quantify and predict freight movements between origin-destination pairs is critical to determining future freight movements by various modes. With the traveling public and policy makers concerned about the congestion, safety, and pollution implications of freight movement, it is imperative that models of freight flows be developed and applied to assist in freight transportation planning efforts.

Major advances have been made in the development of freight transportation modeling methods and frameworks [1-2]. However, many of these methods have not seen application in practice, partially due to the lack of adequate data to support their estimation and application to forecasting. Freight transportation data has been traditionally difficult to collect due to the proprietary nature of the data and due to the difficulty with identifying the proper entity to which a freight transportation survey needs to be administered. The absence of freight transportation data is particularly critical at the disaggregate (spatial and temporal) level, making the estimation of disaggregate models of freight transportation demand a challenge that needs to be addressed. While there are aggregate level freight transportation data sets such as the commodity flow survey (CFS) data, these data sets are generally insufficient to develop models that can estimate origin-destination freight flows by mode and commodity.

The objective of this paper is to describe the development of a structural equations modeling framework for estimating origin-destination freight flows using the Reebie Transearch database. The year Reebie Transearch database used in this paper provided freight flow information for by mode and commodity at the zip code level for the state of Florida. This data was combined with population data from the 2000 Census and with employment data available from another commercial source, InfoUSA, to form a comprehensive database for freight transportation modeling. All databases were available for the year 2000, thus ensuring consistency across databases. A structural equations model that can quantify the freight transportation movement between an origin-destination pair by commodity and mode is estimated and presented. The modeling framework provides a method for estimating freight flows by commodity and mode between origin-destination pairs. Although the model framework was estimated at the zip code level, it may be applicable to a range of spatial levels of aggregation (traffic analysis zone, census tract, cities, counties, etc.). However, care must be exercised before applying the model at a level of aggregation different from that at which model estimation was accomplished.

This paper is organized as follows. The next section provides a brief overview of the model construct and relates it to the previous literature on freight transportation modeling. The third section describes the modeling methodology. Section 4 provides a description of the database used in the study. Section 5 presents sample model estimation results. Finally, concluding remarks are provided in Section 6.

2. MODEL CONSTRUCT

Considerable progress in the estimation and application of freight transportation demand models has been made over the past few decades. It is not possible to offer a comprehensive literature review within the scope of this paper. The significance of freight movement and activity is increasing in terms of both its role in the economy and its potentially adverse impacts on safety and congestion on the transportation system [3-6]. It is not surprising that freight transportation planning has attracted the attention of researchers since the early 1960s [7-9].

Freight travel demand and supply are key elements of the overall freight transportation planning process that also considers the socio-economic environment, intermodal transportation network, policy and regulatory environment, and system performance measures. Some of the models are simple growth factor models while others are more complex and accurate autoregressive integrated moving average models (ARIMA), elasticity models, network models of logistics, direct demand and aggregate demand models, disaggregate demand models, and economic input-output models. Freight transportation models can be used to support a host of planning applications including facility planning, corridor planning, strategic planning, business logistics planning, and economic development. In the area of urban freight travel demand analysis, various truck trip generation methods have been developed [e.g., 10-15]

Freight planning and logistics is concerned with the efficient flow of raw materials, of work in process inventory, and of finished goods from supplier to customer [3]. As the demand for freight transportation and logistics services increases and regional economies continue to grow, new challenges will be faced including the impact of road pricing and freight mobility strategies on freight transportation, the impact of information technology on goods movement, and the role of new developments in business logistics and supply-chain management [3].

Network models of logistics constitute freight network equilibrium models that attempt to model shipper and carrier decisions through a market clearing process [16-20]. Network models that focus on strategic freight network planning employed primarily to forecast, months or years into the future, freight traffic over specific network links and routes and through specific network nodes and terminals have been developed [20]. Work has also been done in the development and application of integrated digital representation of multimodal and transcontinental freight transportation networks for freight transportation modeling [21].

Several policy studies have looked at the role of intermodalism, changing customer requirements, and competition of supply chains in a global marketplace [22-25]. Studies have also been done in the area of potential effects of research and technology on urban freight movement [5] and the need to educate policy makers and transportation planners for the purpose of developing national freight transportation policies. Work has been done to present a framework for organizing and identifying planning goals, key issues, and predominant commodities for intercity freight transportation [22].

The model construct adopted in this paper is largely in line with paradigms and freight transportation demand-supply relationships identified in the literature [25-28]. Figure 1 shows the overall model framework that guided the model development effort of this study. Origin and destination population and employment characteristics are assumed to influence the total amount of flow of a commodity between an origin-destination pair. Modal level of service characteristics including travel distance, travel time, and travel cost influence the total flow between an origin-destination pair and the amount of flow by mode. Thus, this model framework represents a direct demand model where the movement of a commodity by a certain mode is estimated directly from socio-economic characteristics of the origin and destination and the modal level of service variables. Finally, a link is provided from the total flow to the modal flows to accommodate any influence that total freight flow has on individual modal flows. Thus the model framework provides a mechanism by which freight flows can be estimated and changes in freight flows can be determined in response to changes in the socio-economic characteristics of the origin or destination and in response to changes in modal attributes. The model framework is simple and practical and can therefore be easily estimated on a database that can be assembled by any public agency that has resources to purchase some commercial databases.

3. MODELING METHODOLOGY

The modeling methodology adopted in this paper is centered around the structural equations modeling framework that can be used to determine and model relationships among several dependent (endogenous) variables simultaneously. As the model framework described in the previous section includes a number of endogenous variables (freight movements by mode), it was considered appropriate to adopt this modeling methodology. However, in developing the models, it was found that there are many origin-destination pairs that do not exchange freight flows of a certain commodity at all. Thus, there is a high number of zero flows in the database. As the presence of zeros unduly skews the dependent variable distribution (a spike at zero in the freight flow distribution), this paper employs a structural equations estimation methodology that accommodates skewed non-normal endogenous variables.

As mentioned earlier, the modeling of multimodal freight movements involves dealing with multiple endogenous variables in a simultaneous equations framework. Commodity flows on different modes are travel demand related endogenous variables that are inter-related with one another. When modeling the interactions among several inter-dependent endogenous variables, simultaneous equations systems offer an appropriate framework for model development and hypothesis testing [29]. In this paper, the structural equations methodology is adopted for estimating simultaneous equations systems that capture the inter-dependencies among multimodal freight movements.

A typical structural equations model (with 'G' number of endogenous variables) is defined by a matrix equation system as shown in Equation 1.

$$\begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_G \end{bmatrix} = \begin{bmatrix} Y & X \end{bmatrix} \begin{bmatrix} B \\ \Gamma \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_G \end{bmatrix} \tag{1}$$

This can be rewritten as,

$$Y = BY + \Gamma X + \varepsilon \tag{2}$$

(or) $$Y = (I - B)^{-1}(\Gamma X + \varepsilon) \tag{3}$$

where Y is a column vector of endogenous variables,
B is a matrix of parameters associated with right-hand-side endogenous variables,
X is a column vector of exogenous variables,
Γ is a matrix of parameters associated with exogenous variables, and
ε is a column vector of error terms associated with the endogenous variables.

Structural equations systems are estimated by covariance-based structural analysis, also called method of moments. In this methodology, instead of minimizing the sum of squared differences between observed and predicted individual values, the difference between the sample covariances and the covariances predicted by the model is minimized. The observed covariances minus the predicted covariances form the residuals. The fundamental hypothesis for the covariances-based estimation procedures is that the covariance matrix of the observed variables is a function of a set of parameters as shown in Equation 4:

$$\Sigma = \Sigma(\theta) \tag{4}$$

where Σ is the population covariance matrix of observed variables,
θ is a vector that contains the model parameters, and
$\Sigma(\theta)$ is the covariance matrix written as a function of θ.

Equation 4 implies that each element of the covariance matrix is a function of one or more model parameters. The relation of Σ to $\Sigma(\theta)$ is basic to an understanding of identification, estimation, and assessments of model fit. The matrix $\Sigma(\theta)$ has three components, namely, the covariance matrix of Y, the covariance matrix of X with Y, and the covariance matrix of X.

Let Φ = covariance matrix of X, and Ψ = covariance matrix of ε. Then it can be shown that [29]:

$$\Sigma(\theta) = \begin{bmatrix} (I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)^{-1'} & (I - B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I - B)^{-1'} & \Phi \end{bmatrix} \tag{5}$$

Before estimating model parameters, it is first necessary to ensure that the model is identified. Model identification in simultaneous structural equations systems is concerned with the ability to obtain unique estimates of the structural parameters. The identification problem is typically resolved by using theoretical knowledge of the phenomenon under investigation to place restrictions on model parameters. The restrictions usually employed are zero restrictions where selected endogenous variables and certain exogenous variables do not appear on the right hand side of certain equations and selected error correlations are specified to be zero. There are several rules that can be used to check whether a structural equation model system is identified. Detailed discussions on these identification rules may be found in Bollen [29], Johnston and DiNardo [30], and Judge et al [31].

The unknown parameters in B, Γ, Φ, and Ψ are estimated so that the implied covariance matrix, $\hat{\Sigma}$, is as close as possible to the sample covariance matrix, S. In order to achieve this, a fitting function $F(S, \Sigma(\theta))$ which is to be minimized is defined. The fitting function has the properties of being scalar, greater than or equal to zero if and only if $\Sigma(\theta) = S$, and continuous in S and $\Sigma(\theta)$.

Available methods for parameter estimation include maximum likelihood (ML), unweighted least squares (ULS), generalized least squares (GLS), scale free least squares (SLS), and asymptotically distribution-free weighted least squares (ADF-WLS). Each of these methods minimizes the fitting function and leads to consistent estimators of θ. The ADF-WLS method of estimation was used to estimate parameters of structural equations models as the univariate distributions of the endogenous variables are non-normal in that there are substantial numbers of observations for each variable with zero value, which denotes no commodity flow between a zip code pair. For such distributions, the ML coefficient estimates will be consistent, but the estimates of parameter standard errors and the overall model χ^2 goodness-of-fit will likely be biased [32]. Unbiased estimates of standard errors and goodness-of-fit can be generated using the ADF-WLS method [32].

The ADF-WLS estimation method proceeds in three distinct steps. First, it is assumed that each observed endogenous variable is generated by an unobserved normally distributed latent variable. If the latent variable is greater than a censoring level, it is observed; otherwise the censoring level is observed. Each latent variable is assumed to be conditional on the other variables in the system. The problem is to determine the conditional unknown mean and variance of each censored latent variable. This can be done using the Tobit model. An appropriate maximum likelihood estimation procedure for the Tobit model is described in Maddala [33]. Second, estimates of the correlations between the latent censored endogenous variables, and the correlations between each of the latent variables and the continuous exogenous variables in the system are derived. Finally, parameters of the structural equation model are estimated such that the model-implied correlation matrix is as close as possible to the sample correlation matrix, where the sample correlation matrix is determined in the previous steps. The fitting function is then:

$$F_{WLS} = [s - \sigma(\theta)]' W^{-1}[s - \sigma(\theta)] \tag{6}$$

where s is a vector of censored correlation coefficients for all pairs of endogenous and exogenous variables, $\sigma(\theta)$ is a vector of model-implied correlations for the same variable pairs, and W is a positive-definite weight matrix. Minimizing F_{WLS} implies that the parameter estimates are those that minimize the weighted sum of squared deviations of s from $\sigma(\theta)$. This is analogous to weighted least squares re gression, but here the observed and predicted values are variances and covariances rather than raw observations. The best choice of the weight matrix is a consistent estimator of the asymptotic covariance matrix of s:

$$W = ACOV(s_{ij}, s_{gh}) \tag{7}$$

Under very general conditions:

$$W = \frac{1}{N}(s_{ijgh} - s_{ij}s_{gh}) \tag{8}$$

is a consistent estimator, where s_{ijgh} denotes the fourth-order moments of the variables around their means, and s_{ij} and s_{gh} denote covariances. Browne [34] demonstrated that F_{WLS} with such a weight matrix will yield consistent estimates, which are asymptotically efficient with correct parameter test statistics. These properties hold for very general conditions, and consequently such estimators are known as arbitrary distribution function, or asymptotically distribution free (ADF) estimators. ADF-WLS estimators are available in several structural equation model estimation packages including AMOS [35] and LISREL [36].

4. DATABASE PREPARATION AND DESCRIPTION

Two widely known databases for freight transportation planning are the Bureau of Transportation Statistics Commodity Flow Survey (CFS) and the Reebie Associates' TRANSEARCH database. While the former is a free publicly available database, the latter is a commercially available database that may be customized to fit the needs of the planning agency. The Commodity Flow Survey data provides information on commodities shipped, their value, weight and mode of transportation, as well as the origin and destination of shipments at the national, state, and large metro-area levels. Thus, it is quite a detailed aggregate level dataset that can be used to study overall trends in commodity flows between major geographic areas. Thus, it provides a convenient mechanism to obtain control totals regarding freight movements.

In this study, as the intent is to model freight movements at the more microscopic geographic level, the aggregate CFS data is not appropriate. The Florida Department of Transportation, as part of its ongoing research into the development of urban and statewide freight transportation models, purchased the TRANSEARCH freight flow database at the zip code level. The TRANSEARCH data may also be prepared and purchased at other levels of aggregation; however, the zip code level was considered an appropriate level of disaggregation where the data could be considered reliable and large amounts of missing data could be avoided. Thus, the TRANSEARCH 2000 database for the state of Florida was used as the primary source of commodity flow information for this study. In addition, the TRANSEARCH database consists of commodity flows (by ton) into, out of, and through Florida. However, in order to keep the model estimation database tractable, only those commodity flows that originated and ended at zip codes in Florida were used for model estimation. Thus, the database contained commodity flows at the zip code to zip code level with commodities classified at the level of the two-digit Standard Transportation Commodity Classification (STCC) code. In order to reduce the commodity groups to a more manageable level, the commodities at the two digit STCC level were collapsed into 17 commodity groups [37]. The 17 commodity groups are shown in Table 1. It should be noted that the TRANSEARCH database is not necessarily a complete and comprehensive coverage of all freight transportation flows. There are certain types of movements that are not captured in the TRANSEARCH database. Moreover, zip code level data was purchased by the Florida Department of Transportation for all pairs of zip codes that accounted for about 70 percent of the total freight flow. Remaining freight flow data was available only at the county level. This study utilizes only the zip code level data for model estimation and thus captures only about 70 percent of the intra-state freight flow within Florida.

Commodity flows were broadly assigned to four modes: truck, rail, water, and air. The truck mode was further subdivided into full truck load, less than truck load, and private truck load. Full truck load was defined as a “for-hire” commodity flow on a truck with greater than 10,000 pounds, and less than truck load was defined as “for-hire” commodity flow on a truck with less than 10,000 pounds. A Private Truck was defined as any company truck that is part of a private fleet. The Rail mode was also subdivided into rail intermodal and rail car load modes. Thus, the TRANSEARCH dataset consists of freight flows between pairs of zip codes in Florida by commodity (two-digit STCC code) and in annual weight (in tons) for each of the following modes of transport: full-truck-load, less-than-truckload, private truck, rail carload, rail intermodal, water, and air.

In order to model freight flows between zip codes, three more pieces of information were required as per the model construct presented in Figure 1. Socio-economic information represented by population and employment characteristics was needed. All population information was derived from the 2000 Census databases. Census data was obtained at the zip code level and appropriately matched to the TRANSEARCH commodity flow database so that each record contained the population characteristics of the origin zip code and the destination zip code. Employment characteristics were derived from the InfoUSA 2000 database, a commercial database that contains information on every employer in the state of Florida. This database was purchased by the Florida Department of Transportation for use by public agencies in developing employment characteristics for their travel demand models. Aggregation of the employer database was performed to the zip code level and information on employment by SIC category was matched into the TRANSEARCH freight database to develop a comprehensive database with freight flow information, population information, and employment information.

Finally, as per the model construct, detailed information on modal level of service variables is needed. For every zip code pair, it would be ideal to have travel time, distance, and cost information by all modes identified in the database. This effort is currently ongoing and as such all modal level of service variables have not yet been merged into the database. Thus, at this time, the models are estimated using simple map distance (center of zip code to center of zip code) as a measure of impedance between them. However, in the future, modal level of service variables will be included in the model specification to make sure that the model is sensitive to modal level of service attributes. In its current form (as presented in this paper), the model is not sensitive to modal level of service variables. The importance of the role of distance and trip length in freight transportation modeling is well recognized [26, 39].

The distribution of total annual flow (weight) by commodity group is shown in Table 1. As the database is focusing heavily on intra-state movements, the warehousing commodity group is found to account for more than 50 percent of the flows by weight. Other major commodity groups include other minerals, food, and clay, concrete, and glass. The overall mode share of total annual commodity flow by weight is also shown in Table 1. The truck mode accounts for the major portion of overall freight flow, carrying 78 percent of the total annual commodity flows. The truck mode is a combination of two types of trucks, for-hire and private truck. For-hire truck accounts for 37 percent of the total annual commodity flow and about 48 percent of the total commodity flow carried by truck. Private truck accounts for 40 percent of the total annual commodity flow and 52 percent of the total commodity flow carried by truck. Rail mode accounts for about 20 percent of the total annual commodity flow. Air accounts for a very small share of commodity flow by weight at less than one percent while water accounts for a slightly higher share at 2.8 percent.

Table 1 also presents mode shares by commodity group. The table shows the 17 commodity groups and the mode share for each commodity in percent by weight. Thus coal is carried completely by rail car while warehousing is completely moved by truck. Differences across commodity groups with respect to modal share are quite important and noticeable. Therefore, in this study, structural equations models of freight flow are estimated separately for each commodity group. As the commodities vary with respect to density, value, and time-sensitivity, there may be fundamental differences in the relationships among variables that can be used to predict their flows.

Additional exploratory analysis conducted on the database suggested that the database offered variables with plausible statistical distributions and summaries consistent with expectations. Although the TRANSEARCH freight database has its share of errors and omissions, many states are investing in the purchase of this data to develop statewide freight travel demand models. In this context, it was felt that it is not inappropriate to develop structural equations models of freight flow using the TRANSEARCH databases as the objective of this paper is to develop practical models of freight flow that utilize data available at many state and local agencies. However, readers should note the potential limitations of the database used in this study and interpret model results presented in the next section with appropriate caution.

5. MODEL ESTIMATION RESULTS

This section describes the model specification and estimation results for the structural equations models. The models employed a host of exogenous (explanatory) and endogenous (dependent) variables to model freight flows by zip code pair. Exogenous variables may be divided into three groups: population demographic characteristics of the origin and destination, employment characteristics of the origin and destination, and the impedance (distance) between the origin and destination. All population demographic variables were derived from the 2000 Census and all employment characteristics were derived from the InfoUSA 2000 database. Exogenous variables to be included in the models were selected based on earlier research [26, 27]. Endogenous variables are commodity flows between origin-destination zip codes by mode. The commodity flow on each mode is a different endogenous variable. As mentioned earlier, the distributions of the endogenous variables are highly skewed and non-normal with a large number of zero observations. Nearly 97 percent of the observations are zero observations in the data set. Even within the context of the ADF-WLS estimation method, such a heavily zero-inflated distribution leads to computational intractability. To help with computational tractability, log transformations of the variables are used in the estimation process. For all observations and variables, unity was added to the raw variable value to avoid having to take the logarithm of zero which is undefined. Thus all zero observations appear as zeros in the log-transformed data set as well because the logarithm of unity is zero.

In this study, models were estimated for all commodity groups. However, for the sake of brevity, only two types of models are presented in this paper to illustrate the model specification and estimation results. First, models are presented for total freight flows by all commodity groups combined. Second, models are presented for the food commodity group. Many of other commodity groups yielded models rather similar to these models. As such, these models may be considered illustrative of the types of the models that can be developed and applied using the database and methods described in this paper.

Structural equations models were estimated on all origin-destination zip code pairs in Florida. Table 2 presents the structural equation model estimation results for all commodities combined while Table 3 presents the structural equation model estimation results for the food commodity group. The path diagram showing the relationships depicted in Table 2 is shown in Figure 2 while that depicted in Table 3 is shown in Figure 3. The models provided excellent goodness-of-fit measures with the χ^2 statistic indicating that the model can not be rejected with a high degree of confidence (95 percent or higher) and with the goodness-of-fit index (GFI) equal to unity. Thus the models are clearly capable of capturing the key relationships influencing freight flows, even within the context of a large database (more than 850,000 records) where endogenous variables are highly skewed, zero-inflated, and non-normal.

The indications provided by the two models are quite consistent with expectations and plausible. The tables show the direct effects, indirect effects, and total effects that constitute relationships among variables. A direct effect is one where a variable directly affects another variable as depicted by a direct arrow linking the two variables in the path diagram. On the other hand, an indirect effect is one where a variable influences another variable through a mediating variable. For example, in Figure 2, one can see that origin employment does not directly affect the total freight movement by rail. However, origin employment affects both total flow and total truck flow. In turn, total flow and total truck flow affect total flow by rail. Thus origin employment affects the total flow by rail through the intermediate variables, total flow and total truck flow. In some cases, a variable may have both a direct and an indirect effect on another variable. Then the total effect is the sum of the direct and indirect effects.

Tables 2 and 3 show that employment, both at the origin and destination end, has a positive impact on all flows by various modes. It can also be seen that distance has a negative impact on all flows by various modes. Further, the total flow affects the total truck and rail flows as expected with coefficients less than one. These coefficients represent how the total flow between an origin-destination pair contributes to the different types of modal flows between an origin-destination pair. Also, as expected, the total flow by truck has a negative effect on the total flow by rail. Distance is found to have a negative impact on freight flows between origin-destination pairs. Once again, this finding is consistent with expectations as distance constitutes a measure of impedance. While there are certainly strategic level decisions regarding facility location and customer clustering that tend to make distance a secondary variable in influencing freight flows, one can not ignore the possibility that distance is correlated with the quantity of freight flow between an origin-destination pair. For the state of Florida, recent Commodity Flow Surveys have indicated that about 60 percent of freight movements by value and 80 percent of freight movements by weight occur within the state. Clearly, distance is playing a major role in shaping the distribution and quantity of freight flows in Florida. In fact, about 70 percent (by weight) of all commodity flows originating in Florida travel less than 100 miles. The distance variable in the models simply reflects this tendency in the freight flow database and is found to offer statistically significant and intuitively plausible coefficients.

A rather surprising finding is that the origin and destination population variables are found to have a negative impact on freight flows in both the models. It was originally expected that population variables would have a positive impact on the quantity of flow. However, estimation results show that population variables are associated with negative coefficients. On the other hand, the employment variables have positive coefficients. Thus, it appears that employment is the key driver of freight flow activity while resident population is not a key driver of statewide freight flow activity. Business establishments, manufacturing and production operations, and other industrial land uses contribute to heavier volumes of freight flow. The presence of a residential population does not necessarily contribute positively to freight flows between origin-destination pairs at a statewide level. Within an urban area context, when one is concerned with movement of goods and services within an urban area, then one may conjecture that both business establishments and residential population contribute positively to truck trip generation. However, within the context of a statewide freight flow analysis where the freight flows are mostly industrial raw goods, residential population is not likely to attract freight trips. Indeed, many industrial sites are located in zip codes with minimal residential population and attract large amounts of freight flow. Thus, it appears that this finding may have some merit in the statewide modeling context. This finding also lends credence to the approach taken by many states and urban areas that try to attract "jobs" to their area to promote economic activity. The notion is that people will then come to where the "jobs" are located. Finally, it should be noted that previous research in the development of statewide freight trip generation models also found negative coefficients associated with the population variables. In a similar piece of work, Brogan, et. al. [37] provide freight trip generation equations (single production and attraction equations by commodity group) estimated on the TRANSEARCH database. In their equations, the population variables are found to have negative coefficients and employment variables are found to have positive coefficients. Thus the models appear to provide very robust indications of the effects of residential population on origin-destination freight flows by commodity and mode.

6. CONCLUSIONS

Freight transportation model development is now a critical component of the overall transportation planning process as urban areas, states, and the nation consider mobility strategies for enhancing the safety and efficiency of freight transportation. This paper was aimed at developing a simple but practical modeling framework for directly estimating freight transportation flows by commodity between origin-destination pairs. The model development is based on the TRANSEARCH freight flow database that is commercially available from Reebie Associates. This database, providing freight flow information at the zip code level, was merged with population information from the Census 2000 database and employment information from the InfoUSA 2000 database. The resulting database constituted a comprehensive database for modeling freight flows between origin-destination pairs. The only missing component in the database is modal level of service attributes that would potentially influence freight flows by mode (by commodity) between origin-destination pairs. The process of merging modal level of service attributes is currently ongoing and will result in further enhancement of the models developed in this paper.

The modeling methodology consists of a structural equations model that can accommodate multiple dependent variables simultaneously. This structural equations model can be applied to all origin-destination zip code pairs in a region. In this model system, explanatory variables representing origin and destination population and employment characteristics and impedance (distance between the origin-destination pair) are included. The models for various commodity groups are found to offer statistically valid indications and plausible interpretations suggesting that these models may be suitable for application in freight transportation demand forecasting applications.

Undoubtedly, further research is important in the area of freight transportation demand modeling. The development of freight databases and collection of freight movement data continues to be a challenge for model development and estimation. There is always a degree of uncertainty regarding the coverage of the database with respect to geography, commodity groups, and modes and regarding the accuracy of the data as one goes to greater levels of spatial detail. The development of freight transportation models is making great strides, but there is some question as to how transferable these models are between geographic contexts and between geographic scales within the same context. How applicable is it to use a model system estimated at the zip code level at another level of aggregation such as census tract or traffic analysis zone? Research into these issues will greatly enhance our ability to develop freight transportation models and estimate freight flows accurately while analyzing the effects of alternative freight mobility strategies and policies.

ACKNOWLEDGEMENT

The authors thank the Florida Department of Transportation for providing funding for this study and for providing the database used for model development and estimation. The authors are grateful to four anonymous reviewers for their useful suggestions and comments that greatly improved the paper.

REFERENCES

1. Este, G. D. Urban Freight Movement Modeling, *Handbook of Transportation Modeling*, Edited by D.A Hensher and K.J. Button, 2002.

2. Regan, A. C., and Garrido, R. A. Modeling Freight Demand and Shipper Behavior: State of the Art Future Directions, Institute of Transportation Studies, University of California, Irvine, 2002.

3. Regan, A. C., Holguin, J. V, Garland, C., and Miles, H. S. Freight Transportation Planning and Logistics, *Transportation in the New Millennium,* 5p, A1B02: Committee on Freight Transportation Planning and Logistics, Washington, DC Transportation Research Board, 2000.

4. DeWitt, W., and Clinger, J. Intermodal Freight Transportation, *Transportation in the New Millennium*, A1B05: Committee on Intermodal Freight Transport, Washington, D.C., Transportation Research Board, 2000.

5. Czerniak, R. J., Lahsene, J. S., and Chatterjee, A. Urban Freight Movement, What Form Will It Take?, *Transportation in the New Millennium*, 7p, A1B07: Committee on Urban Goods Movement, Washington DC, Transportation Research Board, 2000.

6. Pendyala, R.M., V.N. Shankar, and R.G. McCullough. Freight Travel Demand Modeling: Synthesis of Approaches and Development of a Framework. In *Transportation Research Record* 1725, *Journal of the Transportation Research Board*, TRB, National Research Council, Washington, D.C., 2000, pp. 9-16.

7. Allen, B. W. The Demand for Freight Transportation: A Micro Approach. *Transportation Research*, 11, pp. 9-14, 1977.

8. Baumol, W. J., and Vinod, H. D. An Inventory Theoretic Model of Freight Transport Demand. *Management Science*, 16(7), March, pp. 413-421, 1970.

9. Slavin, H. L. Demand for Urban Goods Vehicle Trips. *Transportation Research Record* 591, National Research Council, Washington, D.C., pp. 32-38, 1976.

10. Cambridge Systematics, Inc. *A Guidebook for Forecasting Freight Transportation Demand*. NCHRP Report 388, Transportation Research Board, National Research Council, Washington, D.C, 1997.

11. Cambridge Systematics, Inc., COMSIS Corporation, and University of Wisconsin-Milwaukee. *Quick Response Freight Manual*. Final Report DOT-T-97-10, U.S. Department of Transportation, 1997.

12. Faris, J. M., and Ismart, D. Freight Modeling Techniques for Small and Medium Sized Areas. Sixth National Conference on Transportation Planning for Small and Medium-Sized Communities, Washington State Department of Transportation, 1999.

13. Hancock, K. L., Karthikeyan, S., and Sreekanth, A. *Freight Transportation Planning: Converting Commodity Volumes to Vehicles*, Final Report, pp. 128. University of Massachusetts, Amherst, 2000.

14. Ogden, K. W. Truck Movement and Access in Urban Areas. *Journal of Transportation Engineering*, *ASCE*, Vol. 117, No. 1, 1991, pp. 71-90.

15. List, G. F., and Turnquist, M. A. Estimating Truck Travel Patterns in Urban Areas. In *Transportation Research Record 1430*, TRB, National Research Council, Washington, D. C., pp. 87-93.

16. Crainic, T. Operations Research Models of Intercity Freight Transportation: The Current State and Future Research Issues. *Logistics and Transportation Review*, 23(2), pp. 189-206, 1987.

17. Harker, P. T. The State of the Art in the Predictive Analysis of Freight Transportation Systems. *Transport Reviews* 5(3), pp. 44-62, 1985.

18. Harker, P. T., and Friesz, T. L. Prediction of Intercity Freight Flows. Part I: Theory and Part II: Mathematical Formulations. *Transportation Research B* 20 (2), pp. 139-153, 155-174, 1986.

19. Bronzini, M. S. Evolution of A Multimodal Freight Transportation Network Model. *Proceedings of the Transportation Research Forum,* 21(1), pp. 475-485, 1980.

20. Friesz, T. Strategic Freight Network Planning Models. In: *Handbook Of Transport Modeling,* pp 527-537, Elsevier Science Publishers, 2000.

21. Southworth, F., and Peterson, B.E. International and Intermodal Freight Network Modeling, *Transportation Research C* 8(1), pp 147-166, 2000.

22. Souleyrette, R., Maze, T. H, Strauss, T., Preissig, D., and Smadi, A. G., Freight Planning Typology. *Transportation Research Record* 1613, pp. 12-19, 1998.

23. Strong, C.K., Harrison, R., and Mahmassani, H. S. A Methodology for Determining the Freight Border Transportation Impact of the North American Free Trade Agreement, University of Texas, Center for Transportation Research, Report 1319-4, 1996.

24. Ogden, K.W. *Urban Goods Movement: A Guide To Policy And Planning*. Ashgate, Aldershot, England, 1992.

25. Shankar, V.N. and R.M. Pendyala. Freight Travel Demand Modelling: Econometric Issues in Multi-level Approaches. In D. Hensher, ed., *The Leading Edge of Travel Behaviour Research*, Elsevier Science Publishers, B.V., The Netherlands, Chapter 38, 2001, pp. 629-644.

26. Holguin-Veras, J., and Thorson, E. Trip Length Distributions in Commodity-Based and Trip-Based Freight Demand Modeling. In *Transportation Research Record 1707*, TRB, National Research Council, Washington, D. C., pp. 37-48.

27. Kumar, A. S., and Bhat, C. R. Freight Modal Split Modeling: Conceptual Framework, Model Structure and Data Sources. *Center for Transportation Research*, CTR, 1833-5.

28. Winston, C. The Demand for Freight Transportation: Models and Applications. *Transportation Research Part A*, 17(6), pp. 419-427, 1983.

29. Bollen, K. A. *Structural Equations with Latent Variables*. Wiley, New York, 1989.

30. Johnston, J., and DiNardo, J. *Econometric Methods*. Fourth Edition, Wiley, New York, 1997.

31. Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T.C. *The Theory and Practice of Econometrics*. Second Edition, Wiley, New York, 1985.

32. Golob, T. F, and McNally, M. G. A Model of Activity Participation and Travel Interactions Between Household Heads. *Transportation Research B*, Vol. 31, No. 3, 1997, pp. 177-194.

33. Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge, MA, 1983.

34. Browne, M.W. Asymptotically Distribution Free Methods for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology* 37, pp. 62-83.

35. Arbuckle, J. L. *AMOS Users Guide Version 4.01*. SmallWaters Corporation, Chicago, IL, 2000.

36. Joreskog, K.G. and Sorbom, D. LISREL 8 Users Reference Guide. Scientific Software, Inc., Chicago, IL, 1993.

37. Brogan, J., Stephen, C. B., and Demetsky, M. J. Application of a Statewide Intermodal Freight Planning Methodology. *Virginia Transportation Research Council*, August 2001, *VTRC*, 02-R5.

38. Reebie Associates. Website http://www.reebie.com/images/transearch.asp, 2003.

39. Garrido, R. A., and Mahmassani, H. S. Forecasting Freight Transportation Demand with the Space-Time Multinomial Probit Model. *Transportation Research B*, Vol. 34, 2000, pp. 403-418.

List of Tables

List of Figures

TABLE 1 Distribution of Freight Flows within Florida by Commodity and Mode

Commodity Group		Freight Flow in Transearch Database		Commodity Flow by Mode in Percentage									Total
				Truck			Rail		Total				
Code	Name	Weight (tons)	Percentage	FTL	LTL	PVT	CL	IMX	Air	Water	Truck Total	Rail Total	
1	Agriculture	71,257	0.07	48.40	0.00	17.15	0.00	27.30	7.13	0.00	65.57	27.30	100
2	Coal	149,729	0.15	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	100.00	100
3	Other Minerals	15,149,353	15.09	0.00	0.00	0.00	95.02	0.00	0.00	4.98	0.00	95.02	100
4	Food	4,505,852	4.49	32.82	0.48	63.64	1.85	1.20	0.02	0.00	96.94	3.04	100
5	Non-Durable Manufacturing	844,392	0.84	29.79	2.81	65.21	0.00	0.74	1.18	0.26	97.82	0.74	100
6	Lumber	1,815,565	1.81	45.87	0.81	47.67	5.65	0.00	0.00	0.00	94.35	5.65	100
7	Paper	542,104	0.54	30.82	6.96	54.66	4.62	2.33	0.61	0.00	92.45	6.94	100
8	Chemicals	6,035,141	6.01	55.26	0.52	0.07	43.32	0.75	0.08	0.00	55.85	44.07	100
9	Petroleum	2,396,887	2.39	5.43	0.13	16.30	0.03	0.15	0.01	77.96	21.85	0.18	100
10	Rubber Plastics	80,190	0.08	24.43	6.60	67.87	0.00	0.70	0.00	0.38	98.93	0.70	100
11	Durable Manufacturing	57,924	0.06	25.09	5.52	20.31	0.00	8.70	39.83	0.55	50.92	8.70	100
12	Clay, Concrete, Glass	14,582,327	14.52	30.16	0.07	67.26	1.17	1.31	0.00	0.02	97.50	2.49	100
13	Primary Metals	220,887	0.22	92.16	0.62	0.00	3.71	2.01	0.01	1.48	92.80	5.72	100
14	Fabricated Metal Products	455,010	0.45	42.37	2.07	50.08	0.20	0.11	4.39	0.77	94.53	0.31	100
15	Transportation Equipment	150,573	0.15	86.74	2.79	0.35	4.66	0.00	3.82	1.62	89.89	4.66	100
16	Miscellaneous Freight	1,917,071	1.91	0.00	0.00	0.00	3.78	87.07	2.45	6.69	0.00	90.85	100
17	Warehousing	51,427,628	51.22	48.72	2.05	49.24	0.00	0.00	0.00	0.00	100.00	0.00	100
	Total	100,401,890	100	36.10	1.21	40.25	17.56	2.00	0.12	2.75	77.56	19.56	100

<u>*Note:*</u>
FTL: Full Truck Load
LTL: Less Than Truck Load
PVT: Private Truck Load
CL: Rail Car Load
IMX: Rail Intermodal Load

TABLE 2 Structural Equations Model Estimation Results for All Commodity Groups

Endogenous Variable	Intercept	Effect	Distance	Destination Employment	Destination Population	Origin Employment	Origin Population	Total Flow	Total Truck	Total Rail	Full Truck Load	Private Truck
Total Flow	-0.370	total	-0.012	0.100	-0.025	0.099	-0.026	0.000	0.000	0.000	0.000	0.000
		direct	-0.012	0.100	-0.025	0.099	-0.026	0.000	0.000	0.000	0.000	0.000
		indirect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total Truck Flow	-0.011	total	-0.009	0.098	-0.025	0.097	-0.025	0.978	0.000	0.000	0.000	0.000
		direct	0.003	0.000	0.000	0.000	0.000	0.978	0.000	0.000	0.000	0.000
		indirect	-0.012	0.098	-0.025	0.097	-0.025	0.000	0.000	0.000	0.000	0.000
Total Rail Flow	0.009	total	-0.011	0.018	-0.004	0.019	-0.006	0.192	-1.545	0.000	0.000	0.000
		direct	-0.003	-0.001	0.001	0.000	-0.001	1.702	-1.545	0.000	0.000	0.000
		indirect	-0.008	0.019	-0.005	0.019	-0.005	-1.511	0.000	0.000	0.000	0.000
Full Truck Load	0.003	total	-0.007	0.084	-0.021	0.083	-0.021	0.845	0.886	0.028	0.000	0.000
		direct	0.000	-0.001	0.000	-0.001	0.000	-0.069	0.930	0.028	0.000	0.000
		indirect	-0.007	0.085	-0.021	0.084	-0.021	0.915	-0.044	0.000	0.000	0.000
Private Truck Load	-0.001	total	-0.007	0.090	-0.023	0.089	-0.023	0.899	0.949	0.017	-0.106	0.000
		direct	0.001	-0.001	0.000	0.000	0.000	-0.065	1.074	0.020	-0.106	0.000
		indirect	-0.008	0.091	-0.023	0.090	-0.023	0.964	-0.125	-0.003	0.000	0.000
Rail Car Load	0.001	total	-0.011	0.017	-0.004	0.018	-0.006	0.191	-1.505	0.999	-0.008	-0.006
		direct	0.000	0.000	0.000	0.000	0.000	-0.039	0.052	0.999	-0.008	-0.006
		indirect	-0.011	0.017	-0.004	0.018	-0.006	0.230	-1.558	0.000	0.001	0.000

Note:

N = 859,329; χ^2 *= 1.064 with df = 6; p-value = 0.983; CFI= 1; RMSEA = 0.000*

All Variables Significant at 95% level

All Variables are in Logarithmic Form

TABLE 3 Structural Equations Model Estimation Results for Food Community Group

Endogenous Variable	Intercept	Effect	Distance	Destination Employment	Destination Population	Origin Employment	Origin Population	Total Flow	Total Truck	Total Rail	Full Truck Load	Private Truck
Total Flow	-0.155	total	-0.006	0.047	-0.015	0.043	-0.013	0.000	0.000	0.000	0.000	0.000
		direct	-0.006	0.047	-0.015	0.043	-0.013	0.000	0.000	0.000	0.000	0.000
		indirect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total Truck Flow	-0.001	total	-0.005	0.047	-0.015	0.043	-0.013	0.994	0.000	0.000	0.000	0.000
		direct	0.000	0.000	0.000	0.000	0.000	0.994	0.000	0.000	0.000	0.000
		indirect	-0.006	0.047	-0.015	0.043	-0.012	0.000	0.000	0.000	0.000	0.000
Total Rail Flow	-0.001	total	0.000	0.002	-0.001	0.002	-0.001	0.051	-1.835	0.000	0.000	0.000
		direct	0.001	0.000	0.000	0.000	0.000	1.876	-1.835	0.000	0.000	0.000
		indirect	-0.001	0.002	-0.001	0.002	-0.001	-1.825	0.000	0.000	0.000	0.000
Full Truck Load	0.003	total	-0.004	0.036	-0.011	0.033	-0.010	0.778	0.830	0.064	0.000	0.000
		direct	0.000	-0.001	0.000	-0.001	0.000	-0.168	0.948	0.064	0.000	0.000
		indirect	-0.004	0.037	-0.011	0.034	-0.010	0.946	-0.118	0.000	0.000	0.000
Private Truck Load	0.001	total	-0.005	0.043	-0.013	0.039	-0.011	0.914	0.928	0.026	0.130	0.000
		direct	0.000	0.000	0.000	0.000	0.000	-0.036	0.853	0.018	0.130	0.000
		indirect	-0.005	0.043	-0.013	0.040	-0.011	0.950	0.075	0.008	0.000	0.000
Rail Car Load	0.000	total	0.000	0.002	-0.001	0.002	0.000	0.050	-1.568	1.010	-0.010	-0.014
		direct	0.000	0.000	0.000	0.000	0.000	-0.288	0.307	1.011	-0.009	-0.014
		indirect	0.000	0.002	-0.001	0.002	-0.001	0.337	-1.875	-0.001	-0.002	0.000

Note:

N = 859,329; χ^2 *= 0.619 with df = 8; p-value = 1.000; CFI= 1; RMSEA = 0.000*

All Variables Significant at 95% level

All Variables are in Logarithmic Form

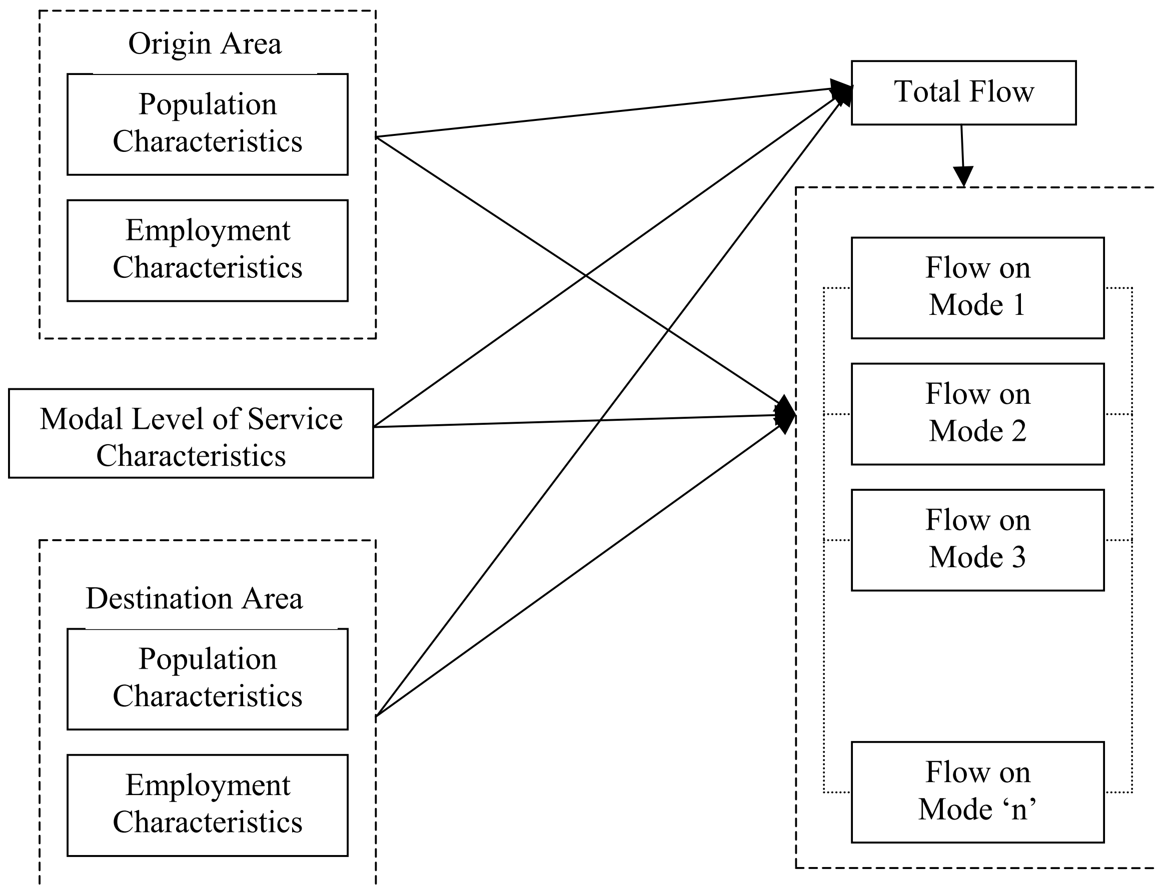

FIGURE 1 Conceptual framework for modeling freight transportation movements.

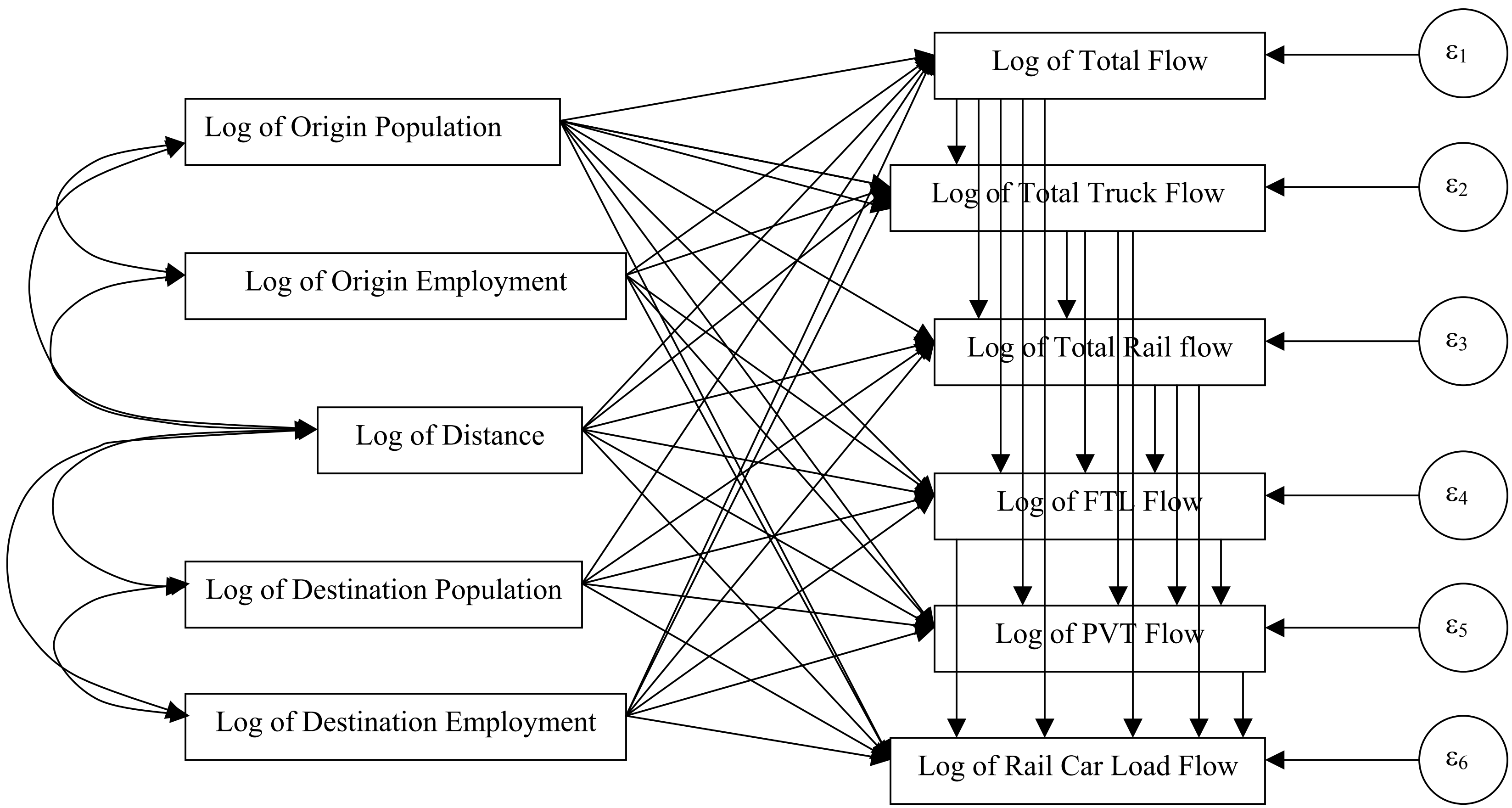

FIGURE 2 Path diagram for the total commodity group structural equations model.

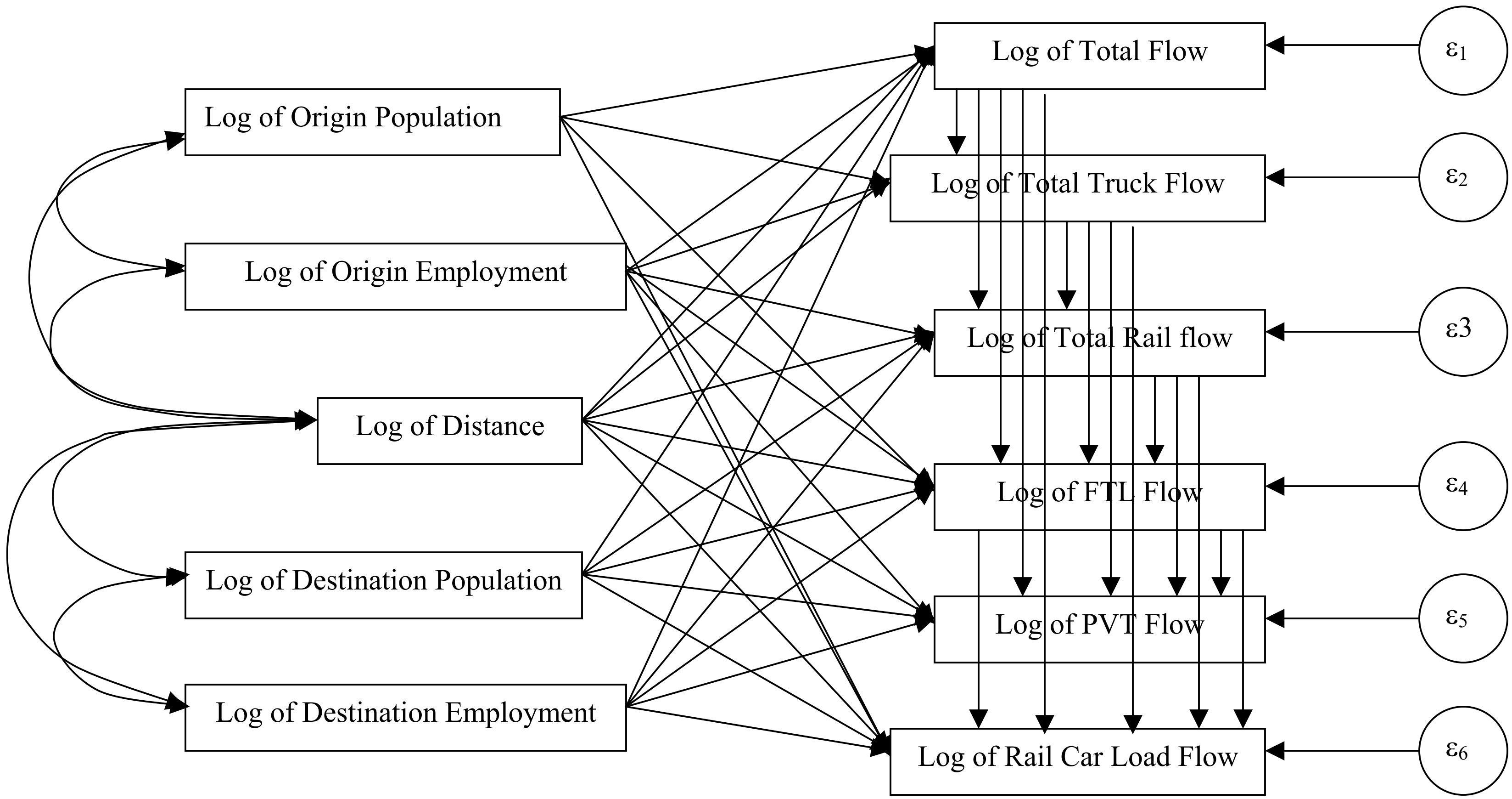

FIGURE 3 Path diagram for the food commodity group structural equations model.