# JOINT FRAMEWORK FOR MODELING FREIGHT MODE AND DESTINATION CHOICE: APPLICATION TO THE US COMMODITY FLOW SURVEY DATA


## ABSTRACT

Earlier research has extensively examined freight mode and shipment weight dimensions. However, freight destination behavior at a high resolution has received scant attention. In our study, we attempt to address the limited research on destination decision processes and develop a latent segmentation-based approach that accommodates mode and destination choices in a unified framework. The proposed approach postulates that these two choices are actually sequential in nature with an infinitesimally small time gap between them. However, the actual sequence (mode-destination or destination-mode) is unknown to us. Thus, a probabilistic model that can accommodate for the two choice sequences within a single framework is proposed. The latent segmentation framework probabilistically assigns the decision maker to the two sequences. In the Mode first – Destination second (MD) sequence, the destination choice model is calibrated with choice alternatives customized to the chosen mode. In the Destination first – Mode second (DM) sequence, the destination model is calibrated without any mode information as mode is unknown to the decision maker. In the study, we used 2012 US Commodity Flow Survey (CFS) data. We found that the latent segmentation-based sequence model outperformed the independent sequence models (MD and DM). The validation exercise also confirmed the superiority of the proposed framework. Finally, an elasticity analysis is conducted to demonstrate the applicability of the proposed model system.



*Keywords:* Freight, mode choice, destination choice, sequence model, latent class model

## INTRODUCTION

In the United States, the contribution of freight to the overall economy has increased substantially over the past half century. With more complex supply chains, growing population, and increasing adoption of online shopping for household goods, it is only likely that the trend will continue. The freight transportation sector contributes to 8.9 percent of the United States' Gross Domestic Product (GDP) while also employing almost 4.5 million people (Freight Facts and Figure, 2017). The Freight Analysis Framework version 4 ($FAF^4$) data predicts that, compared to 2012, freight movements in the US will increase about 42 percent by shipment weight and 87 percent by shipment value by 2040. These expected increases in freight movements will require transportation agencies to pro-actively address infrastructure design and planning related challenges. Towards addressing these challenges, it is important to understand how, where, and how much freight flows. While it would be ideal to model freight movements as a supply chain it is not always feasible to obtain supply chain level data for analysis. Hence, it is common for researchers to examine freight flows as a combination of three decisions. Specifically, study the following dimensions of freight flow: (1) transportation mode (how), (2) destination (where), and (3) shipment weight (how much).

In transportation, several researchers have developed joint models for analysing interconnected decision processes. For example, freight mode and shipment size/weight choice. Generally, these joint models consider transportation mode as a labelled discrete alternative and shipping weight as a continuous or ordered categorical variable (for an exception see Abate et al., 2018). The choice of mode is typically analyzed using a random utility (RU) based model that requires one utility equation per alternative. The shipping weight variable is usually studied using a linear regression (or ordinal regression) model that requires only one propensity equation for all alternatives. For these mathematical representations, researchers develop Simultaneous Equation Models (SEMs) that account for the endogenous nature of the relationship between shipment mode and size choice (common unobserved factors affecting the two dimensions of interest). SEMs can take the form of either a simulation-based approach (for example, see Abdelwahab and Sargious, 1992; Abdelwahab, 1998; ) or an analytically closed form approach (for example, see de Jong and Ben-Akiva, 2007; Windisch et al., 2010; Pourabdollahi et al., 2013; Irannezhad et al., 2019; Keya et al., 2019). The typical dimensionality of the common error terms is obtained as a product of number of equations per choice. In the mode and shipment size case, this is limited to number of mode choice alternatives (as there is only one propensity equation for shipment weight); hence, earlier studies were able to successfully demonstrate these model frameworks. However, the approaches are not directly applicable when one of the dimensions is destination choice.

At an aggregate level, gravity models or input-output models are considered for destination assignment (Ivanova, 2014). At the shipment level (or disaggregate level) destination choice is represented by a large number of alternatives. For example, in the US Commodity Flow Survey (CFS) data, the destination alternative is characterized as a CFS area and there are 132 potential CFS destinations within US for each shipment. The model framework would need to alter from discrete – continuous (or ordered) structure to a discrete – discrete structure where the second discrete model has a large number of alternatives. Thus, adapting the earlier joint frameworks would require us to employ one of two approaches. First, consider the large combinatorial possibilities (number of mode choice alternatives*number of destination choice alternatives) as the choice set. Second, reduce the combinatorial possibilities by parameterizing the destination error correlation as a function of destination attributes. Either approach would require us to resort to simulation-based model developments as closed form solutions become intractable. Simulation based models for a single choice with large number of alternatives is quite cumbersome computationally. Thus, developing joint models with such large choice sets is time consuming and presents with lower prediction accuracy (relative to models with fewer alternatives). To be sure, some research efforts have considered destination choice within a production consumption framework (see de Jong et al., 2017). However, the destinations considered in the models are usually countries and are smaller in number. A similar approach cannot be extended to studies where the potential number of destinations is higher. Given these inherent challenges associated with modeling mode and destination choice behavior, we propose a relatively simpler econometric framework built on earlier work in the passenger demand modeling realm (for example, see Waddell et al., 2007; Chakour and Eluru, 2014; Anowar et al., 2019).

In the proposed approach, we postulate that these joint choices are actually temporally sequential in nature with an infinitesimally small time gap between them. To elaborate, a shipper, while deciding the mode and destination for a shipment, selects one of the choices first. Then, based on the first choice, determines the second one. If the true sequence was known, the joint choice process can be broken down into individual sequential choice process. However, the sequence adopted by the shipper is unknown to us. Thus, in our approach, we recognize that the sequence choice is latent and consider a probabilistic model that can accommodate for all choice sequences within a single framework. The latent segmentation structure allows us to capture the influence of important factors on the sequence decisions while simultaneously modeling shipping mode and destination choices. In our analysis, we allow for two distinct choice hierarchies: (1) mode first – destination second (MD) and (2) destination first – mode second (DM). In this process, the first choice decision is assumed to be known while modeling the second choice decision. This consideration allows us to utilize additional information in modeling the second choice in the sequence. For example, when private truck is the chosen mode, it is more likely that the chosen destination will be within a certain distance (observed from the actual dataset). Hence, for private truck mode in the MD segment, we generate a destination choice set considering the CFS areas which are within 413 miles from the origin ($99^{th}$ percentile routed distance from the dataset). On the other hand, recognizing that the shipment mode is unknown to the shippers in the DM segment and hence, the chosen destination can be any one of the 132 CFS areas in US, we randomly selected 30 unique destinations from the destination choice pool where one of them is the chosen destination CFS area. The reader would note that by not combining mode and destination, the overall number of alternatives is reduced by an order of five (as many modes). This order of magnitude reduction offers computational benefits while also reducing prediction error. The proposed model structure is estimated using 2012 US Commodity Flow Survey (CFS) data. The results clearly highlight the value of the model developed. Additional validation exercises conducted also affirm the benefits of the proposed framework.

The rest of the paper is organized as follows: the following section briefly discusses the existing literature on joint destination and mode choice evaluation in both passenger and freight realms. The third section presents description of the data source and data preparation steps. The section after describes econometric framework used in this study followed by empirical analysis section which discusses model estimation, validation and elasticity results. Finally, the last section concludes the paper.

# EMPIRICAL CONTEXT

## Earlier Research

Mode choice is one of the most extensively researched topics in both passenger travel and freight transportation realm. A review of all the relevant work on freight mode choice behavior is beyond the scope of this paper (see Keya et al., 2017 for a detailed review or Jensen et al. (2019) for recent work). In this section, we focus our attention on research initiatives examining destination choice behavior. From our review, we have observed that earlier literature on the topic can be categorized into two broad groups: (1) studies that examine destination as a separate choice and (2) studies that examine destination in conjunction with mode choice. The majority of the studies in the first group analyze destination choice in the passenger travel context (recreational trip-end (Pozsgay and Bhat, 2001); shopping trip-end (Arentze et al., 2005); tourist destination (Yang et al., 2013); bikeshare trip-end (Faghih-Imani and Eluru, 2015); tour-end (Paleti et al., 2017)), while only a handful of studies examine freight destination choice. We limit ourselves to the discussion of these studies only. We found that all of the studies used stratified importance sampling method (see Ben-Akiva and Lerman, 1985) for creating the destination choice set and used multinomial logit model for analysing the choice behavior (see Mei, 2013; Park et al., 2013; Wang and Holguin-Veras, 2008). The exogenous variables used in these studies include shipping time, distance, number of employment and area type. Of these, travel time/distance is found as the most important factor influencing destination choice.

The second group of studies examine mode and destination choice as a joint decision; again, the majority are from passenger travel behavior context with one exception (Genç et al., 1994). Table 1 provides a summary of these research efforts. The table provides information pertaining to study area, type of mode alternatives, number of destination alternatives, sampling consideration, exogenous variables considered, and methodology employed. Several observations can be made from the table. First, auto (drive alone/passenger), transit, walk, and bike are the most commonly considered mode alternatives. Second, the number of destination alternatives (in the joint decision process) varied from 2 to 1404. Third, methodologically, MNL and nested logit (NL) models are widely used due to their closed form structure and easy interpretability. The reader would note that in MNL and NL models, the mode and destination choices are considered as simultaneous decisions. In a NL model, a pre-determined hierarchy is introduced allowing the consideration of correlation across a subset of alternatives. Within the nesting structure, destination is considered at the upper level and mode is considered at the lower level. More recently, Chakour and Eluru (2014) proposed a latent segmentation-based sequential model for examining commuter train users' access mode and destination station behavior. This is the only study in the literature that allowed for a sequential decision process. Finally, in freight transportation context, Genc et al. (1994) evaluated mode-destination-shipment size jointly using mixed discrete-continuous choice model, where they considered shipment size as continuous variable and mode-destination alternatives as discrete variables. The authors considered only truck and rail as mode alternatives in their study and the number of destination alternatives was limited to 12.

## TABLE 1 Literature on Joint Modeling of Mode and Destination Choice

| Study | Study Area | Decision Variables | | Sampling Considered | Exogenous Variables | Methodology |
|---|---|---|---|---|---|---|
| | | Mode | No. of Destinations | | | |
| **Passenger travel** | | | | | | |
| Richards and Ben-Akiva (1974) | Netherlands | Car, bus, train, moped, walk, bike | 19 | No | Travel time, cost, no. of employment in destination shopping center | Multinomial Logit Model |
| Adler and Ben-Akiva (1976) | Washington D.C., USA | Drive alone, passenger, transit | 134 | No | Travel time and cost, car ownership, distance, no. of retail employment, if destination in CBD, no. of persons in household, household income | Multinomial Logit Model |
| Southworth (1981) | England | Car, transit | 14 | No | Travel time and cost, income, no. of worker in household, distance | Multinomial Logit Model |
| Timmermans (1996) | Netherlands | Car, bus | 2 | No | Travel time, parking cot, travel cost, frequency of bus service, size of shopping center, price level at shopping center, parking facilities, distance | Sequential multinomial logit model |
| Jonnalagadda et al. (2001) | San Francisco, USA | Drive alone, shared ride, transit, walk, bike | 40 | Yes | No. of employment, destination household income, presence of CBD, urban/suburban area, distance, travel time, waiting time, no. of stops, vehicle ownership, no. of worker at household, destination topology, network connectivity, vitality of neighborhood | Nested logit (mode), Multinomial logit (destination) |
| Limanond and Niemeier (2003) | Washington, USA | Auto, bus, walk | 5 | No | Travel time and cost, no. of retail employment in destination, household income, day of week, distance | Multinomial logit model |
| LaMondia et al. (2008) | Europe | Car, air, surface public transport | 6 | No | Home country/abroad, distance, travel companions, age, household size, income, employment status, student, travel planning characteristics, cost at destination, quality of facilities at destination, easily accessible from home, population density, no. of large cities, np. Of hotels, climate, activities for children, friends/family lives at destination, familiar with destination language, product available for shopping, national park/spa/coastal area | Multinomial logit model |

| Study | Study Area | Decision Variables | | Sampling Considered | Exogenous Variables | Methodology |
|---|---|---|---|---|---|---|
| | | Mode | No. of Destinations | | | |
| Yagi and Mohammadian (2008) | Jakarta, Indonesia | Drive alone, shared ride, motorcycle, taxi, transit, non-motorized | 11 | No | Travel time, distance, time of the day, presence and location of intermediate stops, household income, household composition, vehicle ownership, age , gender, destination urban area, land use pattern, density of jobs | Nested logit |
| Newman et al. (2010) | Tennessee, USA | Car, transit, school bus, walk, bike | - | - | No. of student in household, presence of seniors, household income, gas price, bus fare, activity diversity, percent of sidewalk in the zone, household vehicle per person, no. of employment, need river and county border crossing, percent of destination zone within 0.5 mile of bus stop | Nested logit model |
| Seyedabrishami and Shafahi (2013) | Iran | Car, transit | 2 | No | Household car ownership, household size, trip purpose, zonal car ownership, distance from home zone to CBD, travel time | MNL, fuzzy decision tree |
| Chakour and Eluru (2014) | Montreal, Canada | Car, passenger, transit, walk, bike | 18 | Yes | Age, gender, vehicle ownership, employment status, time left home, distance to station, parking facilities at station, travel time, land-use | Latent segmentation based sequential MNL-MNL model |
| Fox et al. (2014) | Toronto, Canada | Drive alone, auto passenger, transit, walk | 1404 | No | Travel time, cost, if destination is CBD, distance, car availability, age, gender, no. of employment, | Nested Logit |
| Ding et al. (2014) | Maryland, Washington D.C.,USA | Car, transit, walk, bike | 3 | No | Household size, income, car ownership, gender, age, residential density, employment density, travel time and cost | Multinomial logit, nested logit, cross-nested logit model |
| **Freight transportation** | | | | | | |
| Genc et al. (1994) | USA | Truck, rail | 12 | No | Waiting time, time for loading unloading, time to travel, market boundary | Mixed continuous/discrete choice model |

### Current Study Context

From the review exercise, it is evident that freight mode and destination choices have rarely been studied in a unified single framework. Even within the destination models from the passenger travel domain, only 2 studies developed joint model of mode and destination with destination alternatives being greater than 100. Further, earlier work typically imposed an *a priori* hierarchy of choice structure (mode at the top level and destination at the lower level or vice versa). Subsequently, the model fit across the two hierarchies are compared and the model with the better fit is chosen. However, it's not necessary that all data records follow the same hierarchy. The current study addresses this limitation by developing a latent segmentation-based approach that allows for two choice sequences: (1) mode first – destination second (MD), and (2) destination first – mode second (DM). The mode and destination choice models estimated in the two sequences differ in exogenous attributes as well as choice alternatives. For instance, in the MD sequence, the mode choice model has no information on the actual destination while the destination choice model is calibrated with choice alternatives customized to the chosen mode. On the other hand, in the DM sequence, the destination model is calibrated without any mode information i.e. the same choice set is adopted for all records (universal choice set of 132 alternatives) while the mode choice model is estimated with the knowledge of destination. Thus, actual travel times to the destination and alternative availability at the destination can be incorporated. To be sure, while the universal choice set has 132 alternatives, our destination choice set for model development comprised of up to 30 alternatives carefully selected using random sampling method. Empirically, the research effort allows us to quantify the impact of various independent variables on mode and destination choice while accounting for the potential interrelationship between the two choices.

## EMPIRICAL DATA

### Data Source

The model in the current study is estimated using 2012 US CFS data. CFS is a shipper based commodity survey carried out every five years since 1993 as part of the Economic Census by the US Census Bureau, in partnership with Bureau of Transportation Statistics (BTS). The 2012 Public Use Microdata (PUM) file contains a total of 4,547,661 shipment records from approximately 60,000 responding businesses and industries. In the data, information is provided on type of commodities shipped, their origin and destination, if special handling is required, distance shipped, shipment value and weight, and their mode(s) of shipping. The commodities are classified by Standard Classification of Transported Goods (SCTG) code. A random sample of 15,000 records was carefully drawn from the PUM database to reduce the data processing and model estimation burden. Care was taken to ensure that the mode shares of the extracted sample matched with the weighted mode shares of the original data. From this sample, 5,000 data records were randomly chosen for model estimation and 10,000 records were set aside for validation exercise. The size of the data sample was based on the data preparation challenges for level of service variables to be employed in the mode choice model. The level of service data (including shipping time, shipping cost) for the various CFS data points were not available in the CFS dataset. The authors generated this information separately for each measure by mode using very time-consuming procedures (documented in Keya et al., 2017). Figure 1 and 2 represents the distribution of SCTG commodity types by shipment value and weight. It is interesting to observe that, by shipment weight, stone and non-metallic minerals has the highest share (27%), whereas by shipment value, this commodity has less than 1 percent share. By weight, only 2 percent electronics products were shipped in 2012, but this commodity had highest share by shipment value (28%).

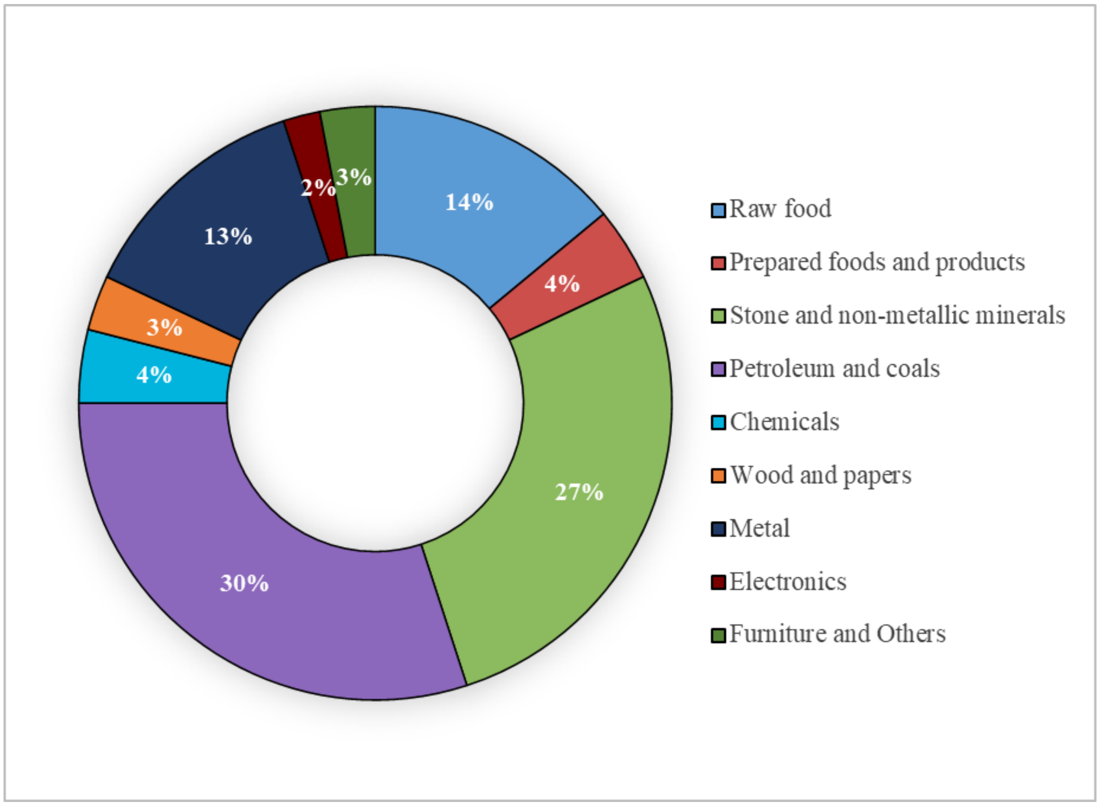


**FIGURE 1 Distribution of SCTG Commodity Type by Shipment Weight**

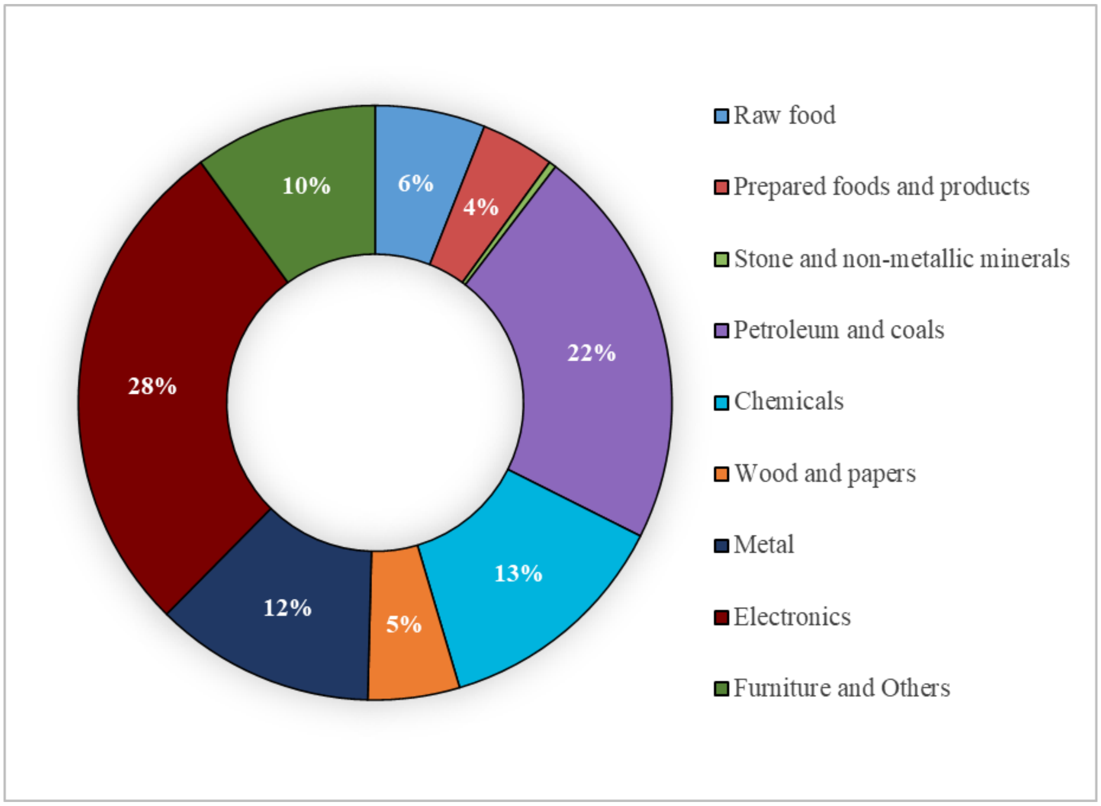


**FIGURE 2 Distribution of SCTG Commodity Type by Shipment Value**

## Data Preparation

*Dependent Variables Generation*

In total, there are twenty-one modes reported in the 2012 CFS PUM file. For our analysis, we consolidated them and created a five-mode category. These include: (1) for-hire truck (including truck and for-hire truck): trucks run by non-governmental business organizations to provide freight transportation facilities to customers under a particular rate; (2) private truck: owned and used by private business units for their own freight movement; (3) air: includes both air and truck mode, as air has limited access and combination of truck with air increases the accessibility; (4) parcel: combination of multiple modes (mainly truck and rail); (5) other mode: involves rail (majority share), water, pipeline or combination of non-parcel multiple modes.

The weighted mode shares in the estimation sample are as follows: for-hire truck (16.71%), private truck (25.55%), air (1.36%), parcel (56.06%), and other mode (0.33%). We adopted a heuristic approach to define mode availability based on observed shipment weight and routed distance. For instance, a 100-ton shipment is very unlikely to be transported by air or parcel mode due to their limitation in carrying capacity and high carrying cost. Private trucks are used for local shipping purposes; hence, they have an intrinsic distance restriction. Keeping these in mind, air and parcel modes are considered available when shipment weight is less than 914 pounds and 131 pounds (99$^{th}$ percentile value from CFS dataset), respectively while private truck is considered available when routed distance is less than 413 miles (99$^{th}$ percentile value from CFS dataset). For-hire truck and other modes are always available (see Keya et al., 2017 for more details).

Destination choice sets for the two sequences are created separately. For DM sequence, the destination choice set comprised of 30 alternatives including the chosen alternative. These 30 alternatives are randomly selected from the 132 available CFS areas. Due to the independence of irrelevant alternative (IID) property of MNL model, the process of random sampling does not affect (bias) the parameter estimates (see McFadden, 1984; Guevara and Ben-Akiva 2013a, Guevara and Ben-Akiva 2013b) and thus has been widely used in destination choice analysis (Pozsgay and Bhat, 2001; Scott et al., 2015; Scott and He, 2012). However, if the same procedure is followed for creating the destination choice set for the MD sequence, it might lead to potential inaccuracy in the empirical analysis. For example, choice of private truck as shipment mode limits possible destination choices; intuitively, destinations within a reasonable distance (as observed from actual data) from the origin would be more preferred. If air or parcel are chosen instead for the same shipment, the destination choice set would invariably expand. To address this issue, new sets of viable destination choice are generated based on chosen mode and its availability. As mentioned before, private trucks are considered available when the routed distance is less than 413 miles. Therefore, for private trucks, the destination choice set comprised of randomly selected CFS areas that are within 413 miles (network distance). Due to the distance restriction, in some instances, we ended up having less than 30 available destination choices for a particular origin. In these cases, we created the choice set using all the available alternatives (including the chosen one) ensuring that all the choice alternatives are unique. For the origins with more than 30 available alternatives, the same methodology previously stated (for creating non-mode specific destination choice sets) is employed. For for-hire truck, air, parcel, and other mode we do not have any shipping distance restrictions. Therefore, for these modes, we randomly chose 30 unique destinations from the 132 CFS areas. Please note that in the mode specific destination choice set, the number of alternatives vary from 5 to 30 (about 21% origins have less than 10 destination alternatives).

*Independent Variables Generation*

We augmented the extracted random sample by level-of-service (LOS) measures, a host of origin-destination (O-D) attributes, and network characteristics. The LOS variables (shipping time and shipping cost) are generated for all available modes. Please note that for different destinations, the shipping time and cost would vary depending on how close/far they are located from the origin. Actual origin and destination location of the shipments is not available in the PUM data – only the CFS area from where the shipment originated and to which the shipment is destined to are provided. Therefore, using network analysis tool in ArcGIS, we generated the network distance from each origin CFS to each destination using their geometric centroids. Using the calculated network distance, we computed the shipping time and cost for each O-D pair employing information from several external sources. The detailed procedure of calculating shipping time and shipping cost is described in Keya et al. (2017). Network and O-D attributes were compiled from various sources including National Transportation Atlas Database (NTAD) 2012, National Bridge Inventory (NBI) data, National Highway Freight Network (NHFN) data, Highway Performance Monitoring System (HPMS) data, Federal Highway Administration (FHWA), and FAF[4] network data. The transportation network attributes include: roadway length per functional classification (interstate highway, freeway and expressway, principal arterial, minor arterial, major and minor collector), railway length, number of airports, number of seaports, number of intermodal facilities, number of bridges, truck Annual Average Daily Traffic (AADT), length of tolled road, length of truck route, length of intermodal connectors, number of truck parking locations, number of truck parking spaces in rest and non-rest areas, ratio of the length of intermodal connectors to the total roadway length, ratio of the length of Primary Highway Freight System (PHFS) and other interstates portions not on PHFS to the total roadway length. The O-D attributes include population density, number of employees, number of establishments by North American Industry Classification System (NAICS) (manufacturing, mining, retail trade, warehouse and storage, company and enterprise, wholesale, information), income categories based on mean income of an area (low (<$50,000), medium ($50,000-$80,000), and high (>$80,000)), number of warehouses and super centers, major industry type in an area (based on the majority of existing industries in a CFS area), percentage of population below poverty level, and annual average temperature (cold if the average annual temperature is less than or equal to 60$^{o}$F; warm if the temperature is greater than 60$^{o}$F) (Weather and Science Facts, 2018).

## METHODOLOGICAL FRAMEWORK

The proposed modeling approach consists of three components: (1) latent segmentation component for the two sequences: (a) mode first-destination second (MD) sequence and, (b) destination first-mode second (DM) sequence, (2) mode choice component for each segment, and (3) destination choice component for each segment. The ensuing presentation is organized by the components.

### Latent Segmentation Component

Let *i* be the index for shippers $(i = 1, 2, \dots, I)$, $q$ be the index for segment $(q = 1\ or\ 2)$, $m$ be the index for mode choice alternative $(m = 1, 2 \dots M)$, and $d$ be the index for destination alternative $(d = 1, 2 \dots D)$. In the latent segmentation component, we determine how shippers are probabilistically assigned to one of the two sequences (MD or DM). The latent process is analyzed using a binary logit structure as follows:

The utility for assigning a shipper $i$ to segment $q$ is defined as:

$$u_{iq}^{*} = \alpha' x_{iq} + \varepsilon_{iq} \tag{1}$$

where $u_{iq}^{*}$ denotes the utility obtained by the $i^{th}$ shipper in selecting the $q^{th}$ segment. $x_{iq}$ is the column vector of attributes which influence the propensity of belonging to segment $q$. $\alpha'$ is the corresponding column vector of coefficients to be estimated. $\varepsilon_{iq}$ is an idiosyncratic error term assumed to follow Type 1 Extreme Value distribution. The shipper $i$ will choose the alternative that offers the highest utility. Then the probability that shipper $i$ belongs to segment $q$ is given as:

$$P_{iq} = \frac{\exp(\alpha' x_{iq})}{\sum_{q=1}^{2} \exp(\alpha' x_{iq})} \tag{2}$$

**Mode and Destination Choice Component**

For mode and destination choice, we employ the random utility based that take the form of two multinomial logit models (see Chakour and Eluru, 2014 for a similar approach). With this notation, the formulation takes the following form:

$$u_{iqm}^{*} = \beta_q' x_{iqm} + \varepsilon_{iqm} \tag{3}$$

$$u_{iqd}^{*} = \gamma_q' x_{iqd} + \varepsilon_{iqd} \tag{4}$$

where $u_{iqm}^{*}$ denotes the utility obtained by choosing mode alternative $m$ in the $q^{th}$ segment, and $u_{iqd}^{*}$ denotes the utility obtained by choosing destination alternative $d$ in the $q^{th}$ segment. $x_{iqm}$, $x_{iqd}$ are column vectors of attributes which influence the choice framework. $\varepsilon_{iqm}$, $\varepsilon_{iqd}$ are idiosyncratic error terms assumed to follow Type 1 Extreme Value distribution. $\beta_q, \gamma_q$ are corresponding column vectors of parameters to be estimated. The second model in each segment is conditional on the first model in the segment. $x_{iqm}$, $x_{iqd}$ incorporate the information available to the shipper at that instant in the choice process.

The probability expression for each model component takes the usual MNL form as follows:

$$P_{iqm} = \frac{\exp(\beta_q' x_{iqm})}{\sum_{m=1}^{M} \exp(\beta_q' x_{iqm})} \tag{5}$$

$$P_{iqd} = \frac{\exp(\gamma_q' x_{iqd})}{\sum_{d=1}^{D} \exp(\gamma_q' x_{iqd})} \tag{6}$$

**Model Estimation**

With these preliminaries, the latent segmentation-based probability for joint choice of mode $m$ and destination $d$ with two segments ($q = 1\ or\ 2$) can be formulated as follows:

$$P_{imd} = P_{i1}P_{i1m}P_{i1d} + P_{i2}P_{i2d}P_{i2m} \tag{7}$$

The first term in Equation (7) reflects the first sequence - MD while the second term reflects the second sequence - DM. The exogenous variables in the second choice for each segment are generated while recognizing the chosen alternative attributes from the first choice process in the segment. The log-likelihood (LL) at the individual shipper level is defined as:

$$L_i = \delta_{md} * ln(P_{imd}) \tag{8}$$

where $\delta_{md}$= 1 if the mode and destination combination is the chosen alternative and 0 otherwise.

$$L = \sum_i L_i \tag{9}$$

The LL function is constructed based on the above probability expression, and maximum likelihood (ML) estimation technique is employed to estimate the $\alpha, \beta_q, \gamma_q$ parameters. The model is programmed in GAUSS matrix programming language. The reader would note that the simplicity of the log-likelihood function does not ensure that the model estimation is straightforward. In fact, the challenges with the estimation of latent models associated with the increase in the degrees of freedom are similar to the empirical identification issues observed in the estimation of simulated maximum likelihood functions for mixed logit models (Cherchi and Guevara 2012). In the simulated maximum likelihood optimization routine, it is very likely that the analyst runs into optimization routine convergence issues because of the flatness of the log-likelihood function. In the latent case, the log-likelihood function in the initial stages of the latent segmentation model is relatively flat thus making it hard to identify the impact of exogenous variables. In fact, some research studies have resorted to the adoption of an EM algorithm for model estimation (Bhat 1997; Kuriyama et al., 2010; Sobhani et al., 2013). In our study, we carefully conduct our model estimation with starting values from the single segment models to alleviate these challenges.

## EMPIRICAL RESULTS

The model specification process was guided by prior research, intuitiveness, and parsimony considerations. We removed statistically insignificant variables (at 80 percent confidence level ($z$-stat = 1.282)) and combined variables when their effects were not significantly different to obtain the final specification.

### Model Fit

We estimated the following models: (1) mode first-destination second (MD) independent model, (2) destination first-mode second (DM) independent model, and (3) latent segmentation-based sequential model (LSS). These models are non-nested; hence, we calculated Bayesian Information Criterion (BIC) for evaluating their performance. The BIC value for a given empirical model can be calculated as: $-2\,(LL) + K\,ln\,(Q)$, where $LL$ is the log-likelihood value at convergence, $K$ is the number of parameters, and $Q$ is the number of observations. The model with the lowest BIC value is the preferred model. The corresponding BIC (number of parameters) values for the MD sequence, DM sequence, and LSS models are: 33,805.96 (21), 31,587.80 (23), and 28,342.14 (43), respectively. The values clearly show that, of the three estimated models, LSS is providing the best data fit. In terms of the two sequences, in this study context, the DM sequence model offers better model fit compared to the MD sequence model.

**Segment Characteristics**

The segment characteristics analysis indicates that, shippers are more likely to choose mode first and then decide the destination location of the shipment (aggregate segment share 53%). In the MD segment, the shipping mode share is as follows: for-hire truck (25.5%), private truck (8.6%), air (3.0%), parcel (62.4%), and other mode (0.5%). In the DM segment, the mode share is as follows: for-hire truck (16.6%), private truck (29.0%), air (0.9%), parcel (51.1%), and other mode (2.4%). These shares clearly illustrate that there is a significant difference in shipment mode shares across the two segments, confirming the presence of heterogeneity. In both segments, parcel mode occupies a larger share. However, the share of for-hire truck is substantially higher in the MD segment, while the share of private truck is significantly higher in the DM segment. The reason may be that when freight destination is decided and is within 413 miles, the shippers are more likely to choose their own vehicle fleet (private truck) for shipping.

**Model Estimation Results**

*Sequence Choice Component*

The second and third columns of Table 2 present the estimates of the latent segmentation component. It examines whether the shippers will choose MD or DM sequence. The positive value of the constant illustrates that when everything remains the same, the probability of choosing MD sequence by the shipper is higher than choosing DM segment. In our analysis, we used freight characteristics as the segmentation variables. We found that when the shipment value is greater than $300, shippers are more inclined to choose MD sequence. One plausible reason might be that when shipping higher value items, shippers are more particular/selective about modes (for example, see Pourabdollahi et al., 2013 and Yang et al., 2014 found that the higher the value of commodity, the higher is the probability of choosing air mode). When the commodity to be shipped is hazardous in nature, shippers are more inclined to choose DM sequence. It is possible that demand for hazardous material is area-specific and depends on the nature of industries in the area. For instance, areas with manufacturing and mining industries might require supply of raw materials that might be hazardous (Pradhananga et al., 2010; Kazantzi et al., 2011; Zhou et al., 2013). Moreover, hazardous materials need careful handling. Once destination is chosen, depending on mode availability and loading-unloading facilities in the destination zone, choosing appropriate mode equipped with the special handling capacity is more convenient. For shipping prepared foods and products, shippers tend to choose destination first and mode later. Market place and residential areas are more likely to attract prepared foods and products. Hence, once destinations are decided, it becomes easier for the shipper to decide on the shipping mode depending on the quantity of product to be shipped. We mapped the probability of belonging to Segment 1 against CFS areas and presented in Figure 3. We can see that, when the origin zones have high density of roadway, railway, airports and ports, shippers are more likely to select the shipping mode first and then destination; presumably, because these transportation features allow better modal accessibility.

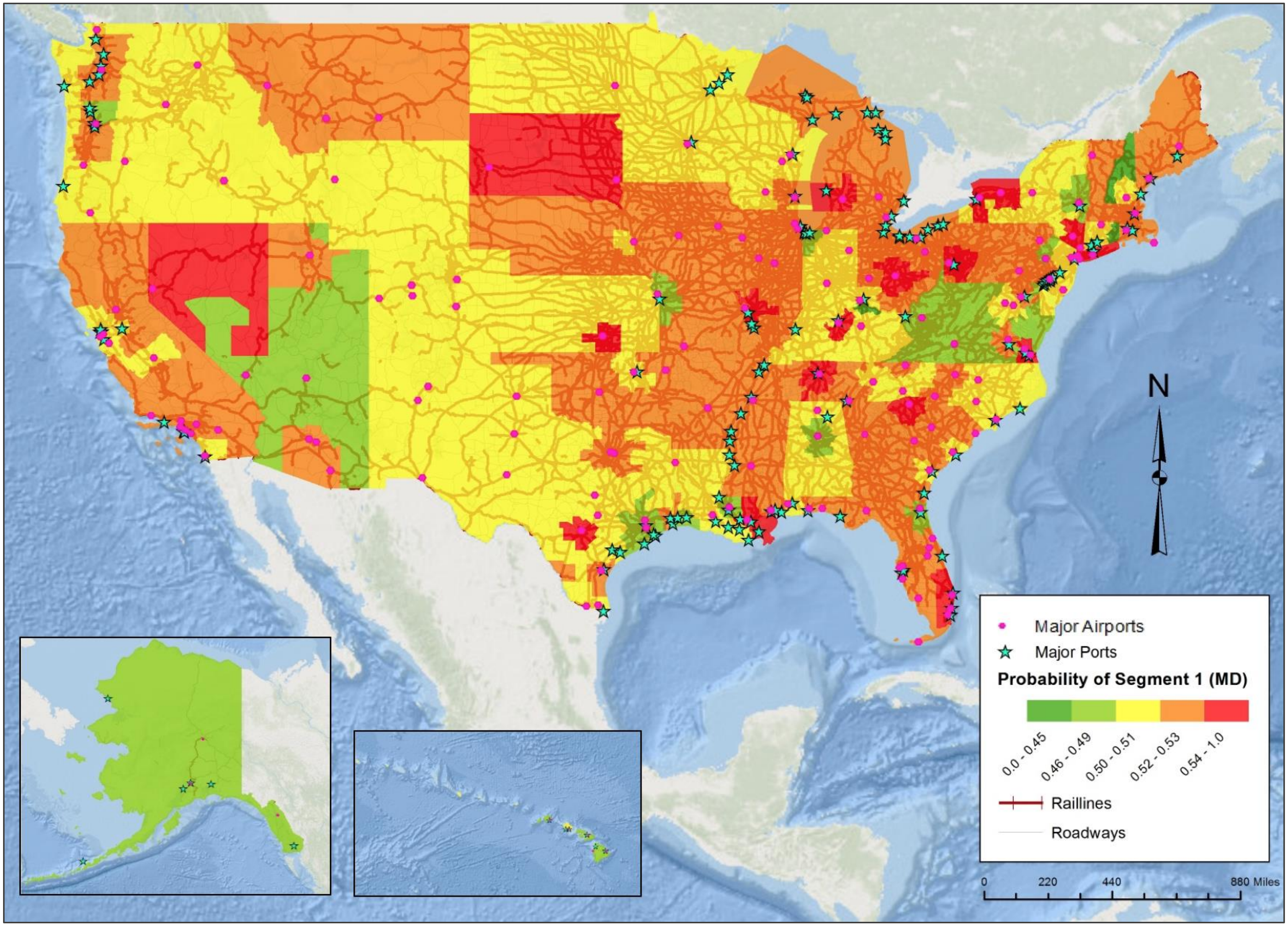


**FIGURE 3 Segment Choice (MD Sequence) Probability Distribution Plot against Origin CFS Areas**

## TABLE 2 Latent Segmentation-Based Mode-Destination Choice Model Results

| **Variables** | **Sequence Choice Results** | | **Shipping Mode Choice Results** | | **Shipment Destination Choice Results** | |
|---|---|---|---|---|---|---|
| | **MD Segment** | **DM Segment** | **MD Segment** | **DM Segment** | **MD Segment** | **DM Segment** |
| | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** |
| Constant | 0.207 (3.709) | - | - | - | - | - |
| Private Truck Constant | - | - | -0.912 (1.704) | 1.768 (13.164) | - | - |
| Air Constant | - | - | -0.69 (-2.343) | -11.525 (-0.074) | - | - |
| Parcel Constant | - | - | 3.18 (12.691) | 3.406 (9.601) | - | - |
| Other Mode Constant | - | - | -3.874 (-11.688) | -27.795 (-3.521) | - | - |
| *Freight Characteristics* | | | | | | |
| Shipment Value: < $ 300 | -0.026 (-1.379) | - | - | - | - | - |
| Hazardous Material | -1.575 (-8.182) | - | - | - | - | - |
| SCTG Commodity Type: Prepared Foods and Products | -1.498 (-8.197) | - | - | - | - | - |
| *Level of Service Variables* | | | | | | |
| Shipping Cost ($1000) | - | - | - | -14.383 (-4.498) | -0.658 (-9.547) | - |
| Shipping Time (100 hrs) | - | - | - | -1.131 (-4.375) | - [1] | - |
| Average Shipping Time To Destination (100 hrs) | - | - | - | - | - | -1.815 (-18.247) |
| *Transportation Network & Demographic Variables* | | | | | | |
| Major Industry in Manufacturing Industry at Origin<br>For-hire Truck | - | - | 0.012 (1.353) | 0.326 (1.973) | - | - |
| No. of Intermodal Facility at Origin<br>Private Truck | - | - | -0.006 (-1.631) | - | | |
| No. of Intermodal Facility at Destination<br>Private Truck | - | - | - | -0.008 (-3.312) | - | - |
| Roadway Density at Origin (mi/ mi$^2$)<br>Parcel | - | - | 0.09 (1.23) | 1.06 (3.237) | - | - |

| Variables | Sequence Choice Results | | Shipping Mode Choice Results | | Shipment Destination Choice Results | |
|---|---|---|---|---|---|---|
| | MD Segment | DM Segment | MD Segment | DM Segment | MD Segment | DM Segment |
| | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** | **Coefficient (t-stat)** |
| Railway Density at Origin (mi/ mi$^2$) Air | - | - | -3.761 (-2.537) | - | - | - |
| Population Density at Origin (1000/mi$^2$) Air | - | - | 0.793 (2.289) | - | - | - |
| Destination Urban Area Air | - | - | - | 9.772 (1.36) | - | - |
| Population Density at Destination (1000/mi$^2$) Air | - | - | - | 0.804 (1.871) | - | - |
| Density of Manufacturing Industry at Destination | - | - | - | - | 0.503 (5.609) | 1.81 (7.807) |
| Density of Management Company and Enterprise at Destination | - | - | - | - | -0.408 (-2.951) | -5.33 (-4.509) |
| No. of Warehouse and Supercenter at Destination | - | - | - | - | 0.014 (17.466) | 0.011 (5.854) |
| Household Income Level at Destination: > $ 80,000 | - | - | - | - | 0.441 (4.973) | 0.935 (4.863) |
| Truck AADT at Destination (million) | - | - | - | - | 0.006 (3.817) | 0.029 (5.391) |
| Density of workers at Destination | - | - | - | - | 1.154 (3.146) | 6.156 (8.452) |
| No. of Truck Parking Location at Destination | - | - | - | - | 1.262 (3.06) | 10.301 (10.588) |
| *Interaction Terms with Chosen Mode* | | | | | | |
| Destination Urban Area*Private Truck | - | - | - | - | -0.707 (-1.466) | - |
| No. of Truck Parking Location at Destination*Private Truck | - | - | - | - | 7.455 (1.294) | - |
| No. of Truck Parking Location at Destination*For-hire Truck | - | - | - | - | 2.295 (3.299) | - |
| No. of Parameters | 43 | | | | | |
| Log-likelihood at Convergence | -13987.95 | | | | | |

[1] = insignificant at 80 percent confidence interval

*Mode-Destination Segment (MD)*

The fourth and sixth columns of Table 2 provides the result of MD sequence. When mode is chosen first, the destination attributes are not known to the shipper. Therefore, destination characteristics or LOS variables which are dependent on the distance from origin to destination, cannot be examined in this sequence. When the major industry type at origin is manufacturing, then the probability of choosing for-hire truck increases. The capacity of carrying larger load from manufacturing industries and better maneuvering flexibility compared to other modes might be a plausible reason (Abdelwahab, 1998; McKinnon, 2006). With increasing number of intermodal facilities at origin, the probability of choosing private truck decreases. Intermodal facilities are intermediaries facilitating the transfer of freight from one mode to another. As private trucks are usually used for shipping within a shorter distance, chances of mode change are lower. Shipments originating from an area with higher highway density is more likely to be shipped by parcel mode as parcel mode requires greater accessibility through roadway network. When the railway density increases at origin, the probability of choosing air mode decreases while increased population density at origin increases the probability.

When destination is chosen second, shipping mode is already known to the shipper. Therefore, LOS variables by mode can be introduced in this segment. As expected, shipping cost negatively impact destination choice; shippers are disinclined to choose a destination for which shipping cost is higher. The result makes intuitive sense. Destinations with higher density of manufacturing industry and increased number of warehouses and supercenters, are more likely to be chosen as they facilitate convenient stockpiling of materials. However, the density of management companies and enterprise negatively influence destination choice. Destination zones with higher household income, truck AADT, density of workers, and number of truck parking locations influence the destination choice positively. The results are intuitive, as higher values of these variables represent higher demand of goods and better facilities for trucking mode. We also tested several interactions between destination attributes and the chosen mode. The results indicate that shippers are less likely to choose destinations in urban area when using private truck as the shipping mode. Operation of truck mode is strictly regulated in urban areas (hence, reducing its accessibility) with high population density; restrictions are imposed on roadway usage (parking, loading, and unloading) along with weight carrying limit. On the other hand, the interactions of number of truck parking locations at destination with for-hire truck and private truck both affect destination choice positively. Understandably, shippers prefer destination areas where they are able to park vehicles easily

*Destination-Mode Segment (DM)*

The fifth and seventh columns of Table 2 present the effects of various variables on DM sequence. When destination is chosen first, the impacts of the variables are quite intuitive. As the mode is unknown to the shipper, the average shipping time for all modes was considered in the model. The impact of average shipping time to destination is found negative as expected. When the density of manufacturing industries and number of warehouses and supercenters at destination increase, the probability of choosing that particular destination increases. Manufacturing industries require raw materials in large quantities to manufacture different products and the number of warehouses and supercenters facilitate storing of raw materials in bulk, attracting more freight flows. However, density of management companies and enterprises affect destination choice negatively. High household income and density of workers increase the attractiveness of the location as freight flow destination. Both variables are indicative of higher demand of goods. Besides, truck AADT and number of truck parking locations at destination also impact destination choice positively; because with increasing truck parking location the accessibility of truck increases.

As destination is already chosen in this segment, therefore shipping time, shipping cost, and destination attributes are known to the shipper and hence, the effects of these variables can be tested in the mode choice model. The negative sign associated with shipping time and shipping cost clearly shows that probability of choosing a particular mode decreases with increasing shipping time and cost. When manufacturing industry is the major industry type at origin, the probability of choosing for-hire truck increases. The shipper is less likely to choose private truck when number of intermodal facilities at destination increases. With increasing roadway density at origin, the probability of choosing parcel mode increases. When destination is an urban area, the probability of choosing air mode increases. The reason may be airports are mainly situated near urban areas. Moreover, with increasing population density at destination, the probability of choosing air mode increases.

**Model Validation**

We also performed a validation exercise using the hold-out sample (10,000 records), in order to ensure that the estimated results are not affected by over-fitting of data. For this purpose, we used the estimated parameters of the final specified models and computed the predictive log-likelihood for the hold-out dataset. The predictive log-likelihood values obtained for MD sequence, DM sequence, and LSS models are: -33548.72, -32372.68, and -28330.62, respectively. The results indicate that the LSS model outperformed the two independent models by a large margin. To further illustrate the improvement in model prediction, we compute predictive log-likelihood values for various sub-samples including Mode, hazardous Material, Export product, Temperature controlled product, shipment value, and destination distance. Table 3 provides the summary of log-likelihood comparison for the three models for the various sub-samples. The results from the table clearly highlight the improvement in model fit for the LSS model across nearly all sub-samples (with small number of exceptions where DM or MD sequence performs slightly better).

The shipping mode shares within the segments are also found similar to the mode shares obtained from the estimation sample. In the MD segment, the shipping mode shares are as follows: for-hire truck (25.6%), private truck (8.4%), air (3.0%), parcel (62.4%), and other mode (0.5%). In the DM segment, the share is as follows: for-hire truck (16.8%), private truck (28.9%), air (0.9%), parcel (51.1%), and other mode (2.3%). Please note that, we followed same data preparation approach for the validation sample as we did for the estimation sample.

**Table 3: Log-likelihood Comparison across Three Models for Different Sub-samples**

| **Variables** | **Sample Share (%), N=10K** | **Log-Likelihood** | | |
|---|---|---|---|---|
| | | **MD sequence** | **DM sequence** | **LSS model** |
| *Mode* | | | | |
| HT | 15.68 | -6435.43 | <u>-5333.08</u> | -5754.43 |
| PT | 24.97 | -8343.36 | -8492.79 | <u>-4876.38</u> |
| Air | 4.09 | -861.39 | <u>-440.98</u> | -860.91 |

| | | | | |
|---|---|---|---|---|
| Parcel | 52.96 | -17641.99 | -18012.74 | -16651.56 |
| Other | 2.30 | -266.56 | -93.09 | -187.34 |
| *Hazardous Material* | | | | |
| Yes | 4.95 | -1546.08 | -1541.18 | -865.95 |
| No | 95.05 | -32002.64 | -30831.50 | -27464.66 |
| *Export* | | | | |
| Yes | 5.46 | -1417.52 | -1352.01 | -1341.07 |
| No | 94.54 | -32131.19 | -31020.66 | -26989.54 |
| *Temperature Controlled Product* | | | | |
| Yes | 4.77 | -1503.98 | -1543.62 | -1186.75 |
| No | 95.23 | -32044.74 | -30829.05 | -27143.87 |
| *Shipment Value* | | | | |
| Shipment Value: <$300 | 42.89 | -15136.65 | -14334.09 | -12783.25 |
| Shipment Value: $300-$1000 | 19.97 | -6816.09 | -6558.78 | -5657.75 |
| Shipment Value: $1000-$5000 | 18.23 | -5928.28 | -5933.40 | -5001.10 |
| Shipment Value: > $5000 | 18.91 | -5667.70 | -5546.41 | -4888.52 |
| *Destination Distance* | | | | |
| <=100 miles | 35.33 | -12337.58 | -11831.16 | -6701.47 |
| > 100 miles and <=500 miles | 23.31 | -8147.32 | -7624.07 | -6589.63 |
| > 500 miles and <=1000 miles | 16.54 | -5297.89 | -5228.17 | -6025.72 |
| >=1000 miles | 24.82 | -7765.94 | -7689.27 | -9013.79 |

**ELASTICITY ANALYSIS**

The estimated results from Table 2 do not directly provide the exact magnitude of the effects of variables on the probability of mode and destination choices. However, it might be possible that the effects of some attributes might differ across the decision choices, particular for different modes. To evaluate this, we compute aggregate level elasticity effects for mode choice only considering a subset of independent variables including travel cost, travel time (for private truck and parcel mode), roadway density at origin, railway density at origin and population density at both origin and destination. In particular, we generated the aggregate change in choice probabilities for each mode in response to the increase of the explanatory variables by 10% (see Eluru and Bhat, 2007; Bhowmik et al., 2019a,b; Kabli et al., 2020 for detail). The reader would note that travel cost and travel time measures were altered for one mode and the corresponding changes in probability across all alternatives are presented. In our presentation, we show the elasticities for travel cost for hire truck and air modes while in terms of travel cost, we present the elasticities for private truck and parcel mode. Further, for the latent model, we estimate the aggregate level elasticities for the overall sample as well as for each segment separately to emphasize policy repercussions based on the most critical contributory factors. For the overall sample, we took the segmentation probabilities into consideration. The elasticity effect across different modes for different variables are presented in Table 4.

**Table 4 Elasticity Values**

| Variables | Segments | Modes | | | | |
|---|---|---|---|---|---|---|
| | | ***Hire Truck*** | ***Private Truck*** | ***Air*** | ***Parcel*** | ***Other*** |
| Travel Cost (Hire Truck) | Segment 1 (MD) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Segment 2 (DM) | -5.136 | 2.019 | 0.259 | 0.020 | 9.037 |
| | Overall | -2.032 | 1.586 | 0.057 | 0.008 | 7.488 |
| Travel Cost (Air) | Segment 1 (MD) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Segment 2 (DM) | 0.099 | 0.022 | -7.156 | 0.080 | 0.000 |
| | Overall | 0.039 | 0.017 | -1.582 | 0.033 | 0.000 |
| Travel Time (Private Truck) | Segment 1 (MD) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Segment 2 (DM) | 0.085 | -0.110 | 0.029 | 0.040 | 0.000 |
| | Overall | 0.034 | -0.086 | 0.006 | 0.016 | 0.000 |
| Travel Time (Parcel) | Segment 1 (MD) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Segment 2 (DM) | 2.558 | 1.888 | 6.263 | -2.214 | 0.000 |
| | Overall | 1.012 | 1.483 | 1.385 | -0.901 | 0.000 |
| Roadway Density at Origin (mi/ mi2) | Segment 1 (MD) | -0.034 | -0.024 | -0.107 | 0.020 | -0.034 |
| | Segment 2 (DM) | -0.826 | -0.620 | -2.212 | 0.725 | 0.000 |
| | Overall | -0.348 | -0.492 | -0.572 | 0.307 | -0.006 |
| Railway Density at Origin (mi/ mi2) | Segment 1 (MD) | 0.311 | 0.280 | -5.482 | 0.089 | 0.311 |
| | Segment 2 (DM) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Overall | 0.188 | 0.060 | -4.269 | 0.053 | 0.053 |
| Population Density at Origin (1000/mi2) | Segment 1 (MD) | -0.154 | -0.146 | 3.225 | -0.065 | -0.154 |
| | Segment 2 (DM) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Overall | -0.093 | -0.031 | 2.512 | -0.039 | -0.026 |
| Population Density at Destination (1000/mi2) | Segment 1 (MD) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Segment 2 (DM) | -0.068 | -0.022 | 7.518 | -0.097 | 0.000 |
| | Overall | -0.027 | -0.017 | 1.663 | -0.040 | 0.000 |

The following observations can be made based on the elasticity effects presented in Table 4. First, from the elasticity effects presented in Table 4, we can clearly see some significant differences in the mode choice probabilities across two segments for some variables which highlights the importance of adopting the latent segmentation-based approach. For instance, due to the 10% increase in the roadway density at origin, the likelihood of choosing hire truck will reduce in both segments (both MD and DM segments). However, the reduction rate is more significant (-0.826 vs -0.034)) if destination is chosen first (DM segment). Second, the results clearly indicate that travel mode shares are sensitive to the level of service attributes (travel cost and travel time) as reported in earlier literature (Rich et al., 2009). Specifically, we observe that increase in travel cost for both hire truck and air mode leads to a drop in the probability of choosing that particular mode. However, the impact is higher for hire truck which suggests that if the cost increases by same amount (10%) for both modes, shippers are less likely to choose the hire truck relative to air mode for their deliveries. In terms of travel time, we find similar trends. Increase in travel time for private truck and parcel mode results in reduced likelihood of choosing that particular mode with higher impact on parcel mode. Third, increased roadway density results in increased likelihood of choosing parcel mode and the share of other four modes reduced. Similarly, with increase in railway density at origin, shippers are more inclined to choose the hire truck and parcel modes whereas the preferences for private truck, air and other modes reduced. Finally, changes in population density both at the origin and destination provides similar results. With higher population density, shippers will switch to air mode as indicated by the positive value in Table 4.

## CONCLUSIONS AND FUTURE WORK

This paper investigates the joint decision of shipping mode and destination choice. Here, two sequences are considered for the two choices: mode first-destination second and destination first-mode second. As analysts, we are not privy to the sequence chosen by the shipper. Hence, we considered a probabilistic approach that accommodates the two sequences in a unified model with two segments with each segment representing one sequence of choice. In this process, the first choice decision is assumed to be known while modeling the second choice decision. This consideration allows us to utilize additional information in modeling the second choice in the sequence. As a result, the proposed latent segmentation framework allows us to capture the influence of important factors on the choice of sequence while simultaneously modeling the shipping mode and destination choices. In the MD sequence, the mode choice model has no information on the actual destination. The destination choice model is calibrated with choice alternatives customized to the chosen mode. In the DM sequence, the destination model is calibrated without any mode information i.e. the same choice set is adopted for all records.

The estimation results from the proposed joint model offers intuitive results. The model fit clearly shows that the latent segmentation-based sequence model performs better than the individual sequence model (MD or DM). The population shares in two segments are different with significant difference in mode shares clearly demonstrating the presence of population heterogeneity. While shipping higher value items, shippers are more inclined toward choosing mode first and then the destination. On the other hand, when shipping hazardous materials or prepared foods and products, they are more inclined to choose destination first as demand of these products depend on the nature of the destination area (industrial or marketplace).

In order to better understand the magnitude of the effects of exogenous variables on the mode choice decision, we compute aggregate level elasticity effects for several variables of interest. We estimate the elasticities for the overall sample as well as for each segment separately to emphasize differences across the two segments. The results from our elasticity analysis provide intuitive policy interpretations. Moreover, from the elasticity effects, we can clearly see some significant differences in the mode choice probabilities across two segments.

The proposed model framework can be employed by freight transportation planners to generate mode and destination outcomes as a function of various freight characteristics in the country. However, given the model complexity and the current state of practise, we consider our model to serve primarily as a framework to identify important variables. The elasticity approach described illustrates the process for identification of important variables influencing the mode and destination choice processes. Finally, certain drawbacks of this study need to be acknowledged. The proposed latent segmentation framework cannot be compared to observed real world data as information on choice sequence is rarely collected in data collection exercises for freight (and even passenger) transportation. Hence, it is not easy to hypothesize variable impacts across the various sequences. Hence, it is important that the models are estimated and interpreted carefully.

US CFS 2012 data does not provide exact geo-coded locations of origin and destination of freight movement; rather, it provides the origin and destination at the CFS area level. Any information of trip chaining or any intermediate location of the trip is also unavailable in the dataset. In future, availability of this kind of information will lead to more accurate and intriguing analysis. Methodologically, several avenues for future research exist. The proposed sequential framework can be compared with several potential model systems such as (a) joint mode and destination based nested formulation with alternative hierarchies (mode at the top level or destination at the top level) and (b) latent segmentation-based model with alternative nesting structures as the two segments. Further, in our analysis, we focused on destination choice. It might also be interesting to explore freight origin as a dependent variable (Samimi et al., 2010; Outwater et al., 2013). The consideration could add a new dependent variable to the model and offer an interesting extension. Also, unobserved heterogeneity can be accommodated within the proposed structure to capture the influence of unobserved factors affecting the choice decisions. The computational complexity of the process will increase substantially and might result in estimation challenges.

## REFERENCES


1. Abate, M., I. Vierth, R. Karlsson, G.C. de Jong and J. Baak (2018). A disaggregate stochastic freight transport model for Sweden, Transportation, https://doi.org/10.1007/s1116-018-9856-9
2. Abdelwahab, W., & Sargious, M. (1992). Modelling the Demand for Freight Transport: A New Approach. *Journal of Transport Economics and Policy*, *26*(1), 49-70.
3. Abdelwahab, W. M. (1998). Elasticities of Mode Choice Probabilities and Market Elasticities of Demand: Evidence from a Simultaneous Mode Choice/Shipment-Size Freight Transport Model. *Transportation Research Part E: Logistics and Transportation Review, 34*(4), 257-266.
4. Adler, T. J., & Ben-Akiva, M. (1976). Joint-Choice Model for Frequency, Destination, and Travel Mode for Shopping Trips. *Transportation Research Record: Journal of the Transportation Research Board*, *569*, 136-150.
5. Anowar, S., Faghih-Imani, A., Miller, E. J., & Eluru, N. (2019). Regret Minimization Based Joint Econometric Model of Mode Choice and Time of Day: A Case Study of University Students in Toronto, Canada. *Transportmetrica A: Transport Science*, *15*(2), 1214-1246.
6. Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. Transportation science, 31(1), 34-48.
7. Bhowmik, T., Yasmin, S., & Eluru, N. (2019a). A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. Analytic methods in accident research, 21, 13-31.
8. Bhowmik, T., Yasmin, S., & Eluru, N. (2019b). Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types. Analytic Methods in Accident Research, 24, 100107.

9. Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*: Cambridge, Mass.: MIT Press, c1985.
10. Chakour, V., & Eluru, N. (2014). Analyzing Commuter Train User Behavior: A Decision Framework for Access Mode and Station Choice. *Transportation, 41*(1), 211-228.
11. Cherchi, E., & Guevara, C. A. (2012). A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance–covariance matrix. Transportation Research Part B: Methodological, 46(2), 321-332.
12. Current Results. Weather and Science facts. Average Annual Temperature for Each US State. (2018). Retrieved from https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php.
13. de Jong, G., & Ben-Akiva, M. (2007). A Micro-Simulation Model of Shipment Size and Transport Chain Choice. *Transportation Research Part B: Methodological, 41*(9), 950-965.
14. de Jong, G., Tanner, R., Rich, J., Thorhauge, M., Nielsen, O.A. & Bates, J. (2017). Modelling Production-Consumption Flows of Goods in Europe: The Trade Model within Transtools3. *Journal of Shipping and Trade, 2*(1), 5.
15. Ding, C., Xie, B., Wang, Y., & Lin, Y. (2014). Modeling the Joint Choice Decisions on Urban Shopping Destination and Travel-to-Shop Mode: A Comparative Study of Different Structures. *Discrete Dynamics in Nature and Society, 2014*, 10 pages.
16. Eluru, N., and Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention, 39 (5), 1037–1049.
17. Faghih-Imani, A., & Eluru, N. (2015). Analysing Bicycle-Sharing System User Destination Choice Preferences: Chicago's Divvy System. *Journal of Transport Geography, 44*, 53-64.
18. Fox, J., Daly, A., Hess, S., & Miller, E. (2014). Temporal Transferability of Models of Mode-Destination Choice for the Greater Toronto and Hamilton Area. *Journal of Transport and Land Use, 7*(2), 41-62.
19. Freight Facts and Figures. (2017). Retrieved from https://www.bts.gov/bts-publications/freight-facts-and-figures/freight-facts-figures-2017-chapter-5-economic.
20. Genc, M., Inaba, F. S., & Wallace, N. E. (1994). From Disaggregate Mode-Destination-Quantity Decisions to Predictions of Aggregate Freight Flows. *International Journal of Transport Economics / Rivista internazionale di economia dei trasporti, 21*(3), 269-285.
21. Guevara, C.A. and M.E. Ben-Akiva (2013a). Sampling of alternatives in logit mixture models, *Transportation Research B, 58,* 31-52.
22. Guevara, C.A. and M.E. Ben-Akiva (2013b). Sampling of alternatives in multivariate extreme value (MEV) models, *Transportation Research B, 58,* 185-198.
23. Irannezhad, E., Prato, C., & Hickman, M. (2019). A Joint Hybrid Model of the Choices of Container Terminals and of Dwell Time. *Transportation Research Part E: Logistics and Transportation Review, 121,* 119-133.
24. Ivanova, O. (2014) Modelling inter-regional freight demand with input-output, gravity and SCGE methodologies, in L. Tavasszy and G.C. de Jong (Eds.): Modelling freight transport, Elsevier Insights Series, Elsevier, London/Waltham.
25. Jensen, A.F., Thorhauge, M., de Jong, G., Rich, J., Dekker, T., Johnson, D., Cabral, M.O., Bates, J. & Nielsen, O.A. (2019). A Disaggregate Freight Transport Chain Choice Model for Europe. *Transportation Research Part E: Logistics and Transportation Review, 121*, 43-62.
26. Jonnalagadda, N., Freedman, J., Davidson, W., & Hunt, J. (2001). Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice

Models. *Transportation Research Record: Journal of the Transportation Research Board, 1777*, 25-35.
27. Kabli, A., Bhowmik, T., & Eluru, N. (2020). A Multivariate Approach For Modeling Driver Injury Severity By Body Region. Analytic Methods in Accident Research, 100129.
28. Kazantzi, V., Kazantzis, N., & Gerogiannis, V. C. (2011). Risk Informed Optimization of a Hazardous Material Multi-Periodic Transportation Model. *Journal of Loss Prevention in the Process Industries, 24*(6), 767-773.
29. Keya, N., Anowar, S., & Eluru, N. (2017). *Estimating a Freight Mode Choice Model: A Case Study of Commodity Flow Survey 2012 (No. 17-03631).* Paper presented at the 96th Annual Meeting of the Transportation Research Board, Washington, D.C.
30. Keya, N., Anowar, S., & Eluru, N. (2019). Joint model of freight mode choice and shipment size: A copula-based random regret minimization framework. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 97-115.
31. Kuriyama, K., Hanemann, W. M., & Hilger, J. R. (2010). A latent segmentation approach to a Kuhn–Tucker model: An application to recreation demand. Journal of Environmental Economics and Management, 60(3), 209-220.
32. LaMondia, J., Snell, T., & Bhat, C. R. (2010). Traveler Behavior and Values Analysis in the Context of Vacation Destination and Travel Mode Choices: European Union Case Study. *Transportation Research Record: Journal of the Transportation Research Board, 2156*, 140-149.
33. Limanond, T., & Niemeier, D. A. (2003). Accessibility and Mode-Destination Choice Decisions: Exploring Travel in Three Neighborhoods in Puget Sound, WA. *Environment and Planning B: Planning and Design*, *30*(2), 219-238.
34. McFadden, D. (1987). Regression-Based Specification Tests for the Multinomial Logit Model. *Journal of Econometrics, 34*(1), 63-82.
35. McKinnon, A. (2006). Life without Trucks: The Impact of a Temporary Disruption of Road Freight Transport on a National Economy. *Journal of Business Logistics, 27*(2), 227-250.
36. Mei, B. (2013). Destination Choice Model for Commercial Vehicle Movements in Metropolitan Area. *Transportation Research Record: Journal of the Transportation Research Board*, *2344*, 126-134.
37. Newman, J. P., & Bernardin, V. L. (2010). Hierarchical Ordering of Nests in a Joint Mode and Destination Choice Model. *Transportation, 37*(4), 677-688.
38. Outwater, M., Smith, C., Wies, K., Yoder, S., Sana, B. and Chen, J. (2013). Tour based and supply chain modeling for freight: integrated model demonstration in Chicago. *Transportation Letters: the International Journal of Transportation Research*, *5*(2): 55-66.
39. Paleti, R., Faghih Imani, A., Eluru, N., Hu, H.-H., & Huang, G. (2017). An Integrated Model of Intensity of Activity Opportunities on Supply Side and Tour Destination & Departure Time Choices on Demand Side. *Journal of Choice Modelling*, *24*, 63-74.
40. Park, H., Park, D., Kim, C., Kim, H., & Park, M. (2012). A Comparative Study on Sampling Strategies for Truck Destination Choice Model: Case of Seoul Metropolitan Area. *Canadian Journal of Civil Engineering*, *40*(1), 19-26.
41. Pourabdollahi, Z., Karimi, B., & Mohammadian, A. (2013). Joint Model of Freight Mode and Shipment Size Choice. *Transportation Research Record: Journal of the Transportation Research Board*, *2378*, 84-91.

42. Pozsgay, M., & Bhat, C. (2001). Destination Choice Modeling for Home-Based Recreational Trips: Analysis and Implications for Land Use, Transportation, and Air Quality Planning. *Transportation Research Record: Journal of the Transportation Research Board*, *1777*, 47-54.
43. Pradhananga, R., Taniguchi, E., & Yamada, T. (2010). Ant Colony System Based Routing and Scheduling for Hazardous Material Transportation. *Procedia - Social and Behavioral Sciences*, *2*(3), 6097-6108.
44. Rich, J., Holmblad, P. M., & Hansen, C. O. (2009). A weighted logit freight mode-choice model. Transportation Research Part E: Logistics and Transportation Review, 45(6), 1006-1019.
45. Richards, M. G., & Ben-Akiva, M. (1974). A Simultaneous Destination and Mode Choice Model for Shopping Trips. *Transportation*, *3*(4), 343-356.
46. Samimi, A., Mohammadian, A., & Kawamura, K. (2010). Freight Demand Microsimulation in the U.S. World Conference on Transport Research (WCTR): Lisbon.
47. Scott, D. M., & He, S. Y. (2012). Modeling Constrained Destination Choice for Shopping: A GIS-Based, Time-Geographic Approach. *Journal of Transport Geography*, *23*, 60-71.
48. Seyedabrishami, S., & Shafahi, Y. (2013). A Joint Model of Destination and Mode Choice for Urban Trips: A Disaggregate Approach. *Transportation Planning and Technology, 36*(8), 703-721.
49. Sobhani, A., Eluru, N., & Faghih-Imani, A. (2013). A latent segmentation based multiple discrete continuous extreme value model. Transportation Research Part B: Methodological, 58, 154-169.
50. Southworth, F. (1981). Calibration of Multinomial Logit Models of Mode and Destination Choice. *Transportation Research Part A: General, 15*(4), 315-325.
51. Theo A. Arentze, Harmen Oppewal, & Timmermans, H. J. P. (2005). A Multipurpose Shopping Trip Model to Assess Retail Agglomeration Effects. *Journal of Marketing Research, 42*(1), 109-115.
52. Timmermans, H. J. P. (1996). A Stated Choice Model of Sequential Mode and Destination Choice Behaviour for Shopping Trips. *Environment and Planning A: Economy and Space, 28*(1), 173-184.
53. Waddell, P., Bhat, C., Eluru, N., Wang, L., & Pendyala, R. M. (2007). Modeling Interdependence in Household Residence and Workplace Choices. *Transportation Research Record*, *2003*, 84-92.
54. Wang, Q., & Holguín-Veras, J. (2008). Investigation of Attributes Determining Trip Chaining Behavior in Hybrid Microsimulation Urban Freight Models. *Transportation Research Record: Journal of the Transportation Research Board, 2066*, 1-8.
55. Windisch, E., De Jong, G., Van Nes, R., & Hoogendoorn, S. (2010). *A Disaggregate Freight Transport Model of Transport Chain and Shipment Size Choice.* Paper presented at the ETC 2010: European Transport Conference, Glasgow, UK, 11-13 October 2010.
56. Yagi, S., & Mohammadian, A. (2008). Joint Models of Home-Based Tour Mode and Destination Choices: Applications to a Developing Country. *Transportation Research Record: Journal of the Transportation Research Board*, *2076*, 29-40.
57. Yang, D., Ong, G. P., & Chin, A. T. H. (2014). An Exploratory Study on the Effect of Trade Data Aggregation on International Freight Mode Choice. *Maritime Policy & Management*, *41*(3), 212-223.

58. Yang, Y., Fik, T., & Zhang, J. (2013). Modeling Sequential Tourist Flows: Where Is the Next Destination? *Annals of Tourism Research*, *43*, 297-320.
59. Yasmin, S., & Eluru, N. (2016). Latent segmentation based count models: analysis of bicycle safety in Montreal and Toronto. Accident Analysis & Prevention, 95, 157-171.
60. Zhou, Z., Chu, F., Che, A., & Zhou, M. (2013). ε-Constraint and Fuzzy Logic-Based Optimization of Hazardous Material Transportation Via Lane Reservation. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 847-857.