March 21-23, 2013

**DOUBLETREE HOTEL
ANNAPOLIS, MARYLAND**

Proceedings of the
54th Annual
Transportation Research Forum

TRF

www.trforum.org

DEMAND FOR FREIGHT TRANSPORTATION IN THE U.S.: A HIGH-LEVEL VIEW

Adrian Albert, Stanford University
Andreas Schaefer, University College London

ABSTRACT

Research on freight transportation has seen a tremendous increase in the last decades, yet it still lags behind that on passenger travel, particularly at a macro-level suitable for nation-wide policy analysis. A key challenge in freight demand modeling is the availability of data on key drivers of demand - such as cost, time, and trip length - which usually is proprietary and expensive. Moreover freight data available to the public is usually heterogeneous and published by a number of different bodies. In this study we integrate many publicly-available datasets on these attributes into a consistent database describing freight movement at the U.S. level. We then use this dataset to estimate a discrete-choice model of the shares of major single modes - truck, rail, and air, and compare our results with other similar exercises from the transportation economics literature. We thus present an analysis of the effect of generalized transportation costs and infrastructure quality - captured by travel time - on modal split at the national level. We conclude with recommendations regarding freight transportation data that arise from the insights offered by this exercise for policy-makers and practitioners.

1. INTRODUCTION AND BACKGROUND

Transportation of goods is a major force that drives economic prosperity and has attracted much interest from researchers over the past 30 years from both methodological and empirical standpoints. Despite its importance, and more critically than in passenger transportation, freight demand modeling suffers from low availability of data on key drivers of demand (such as cost, time, and trip length), both at the micro- and at the macroeconomic level. Freight transportation surveys are expensive to conduct, are commercially-sensitive, and thus are usually proprietary and unavailable to researchers. Moreover U.S.-level freight data available to the public is usually heterogeneous and published by a number of different bodies typically concerned with just a small niche of the entire freight system.

While undoubtedly vital for the economy, the freight transportation network is also a major consumer of energy, accounting for around 28% of final energy used in the transportation sector, which is responsible for almost 30% of CO_2 emissions in the U.S. Authority (2011). In the last several decades both volume of freight transportation (in ton-mileage) and its energy footprint have increased Schipper et al. (2011), while a pronounced shift to faster (and more energy-intensive) modes has been observed in the U.S. and in a number of European countries Nijkamp et al. (1999). However accounting for the relationships between main operational factors of different transportation modes and their energy intensity has only currently come again into the attention of researchers Gucwa and Schafer (2011).

The goal of the present paper is to complement the transportation literature by analyzing the impact of the main drivers of U.S.-level, aggregate freight demand, with a particular emphasis on the effect of cost and infrastructure quality on modal competition. We arrive at a macroscopic picture of the market and logistics factors (which are also connected to energy use, e.g., Gucwa and Schafer (2011)) in the transportation of freight. For this we first perform a comprehensive survey of the econometric literature on freight transportation as described in Section 2. In Section 3 we review a standard discrete-choice model that we used in our estimations. We then identify and integrate publicly-available data on U.S. freight transportation, a process which is detailed in Section 4; this is a main contribution of our research effort. To achieve this we used several national survey datasets to develop high-level models of the market and geographical factors influencing logistics decisions (e.g., speed and cost of transportation). Next we used this dataset to estimate a high-level econometric model of freight demand in Section 5, building upon the current state of the practice; note that most previous models were estimated primarily on data from small-scale disaggregate surveys including just a handful of shippers or on state-level data. We validate our model using literature estimates of elasticities of freight attributes from previous U.S. and international studies. A typical quantity of interest in transportation demand (both freight and passenger) that we also estimate is the value of time (VOT) Massiani (2003), which we use to further validate our results. We conclude in Section 6 with thoughts on what data would be necessary to improve the understanding of nation-wide freight transportation system.

2. ECONOMETRIC FREIGHT TRANSPORTATION LITERATURE

Empirical literature on freight transportation demand has seen a rapid increase over the last several decades, although it has remained less researched than that on passenger transportation. Numerous specialized studies have been performed that isolate particular aspects of the freight demand problem: operational characteristics and commodity flows for particular modes (e.g., state-wide models of truck trips (Sorratini and Robert L. Smith (2000)), survey-based models of state-wide

trucking costs (Levinson et al. (2005))), high-level demand analyzed through macroeconomic techniques such as Input-Output analysis (e.g., Vilain et al. (1999)), or on studying general or spatial price equilibrium models to arrive at microeconomically-sound descriptions of freight flow characteristics (e.g., "generalized cost" Tavasszy et al. (2009)). Our focus is on transportation econometrics research, in particular those that use the discrete choice framework Ben-Akiva and Lerman (1985); thus we shall concentrate here on reviewing econometric models.

2.1. Freight demand modeling

Several modeling econometric approaches have crystallized that depend in great part on the type of data used to inform the models, and the scope and goals of the modeling exercise. Early literature (e.g, Winston (1981, 1982)) distinguishes between aggregate and disaggregate models based on the nature of the data employed for estimation: aggregate models use region-, or nation-wide aggregate (survey) data on commodity flows (or modal shares) to characterize the high-level trends in demand, whereas disaggregate models use survey data on individual shipments made by single freight companies. Essentially disaggregate models move away from the "representative shipper" concept utilized in the aggregate context, and provide a richer picture of the mechanisms behind the decision-making of shipping freight. Among the first applications of interest of these econometric models were studying intermodal competition, analyzing the effect of service quality on demand, and predicting or quantifying the effects of deregulation of the transportation industry in the 1970s Winston (1982). A comprehensive review of the early (up to late 1980s) econometric literature is given in Zlatoper and Austrian (1989), where the authors stress the importance of identifying the appropriate variables, in particular for aggregate data (e.g., how does one approximate service characteristics such as reliability?), and point out that disaggregate models offer better estimates of attribute elasticities than aggregate ones (because they use data at the level of the individual decision-maker). At the same time, aggregate data can be more useful for the regional and national-level analysis that we attempt here, although it obscures the mechanisms and decisions at finer scale. More recent surveys on econometric modeling literature are performed in Winston (1983) and in Clark, Naughton, Proulx, and Thomas (Clark et al.), where the authors specifically compare elasticity estimates across different modes and models computed for the U.S. The disaggregate studies surveyed in Clark, Naughton, Proulx, and Thomas (Clark et al.) are presented from the perspective introduced in Winston (1982), which divides them into two classes: inventory-based models (that adopt the perspective of an inventory manager that has to deal with multiple production decisions) and behavioral models (that are concerned maximizing utility when presented with a limited number of choices). A recent review on behavioral models (both aggregate and disaggregate) can be found in Samimi et al. (2010). One recurrent question of concern in the disaggregate choice modeling literature is the (continuous) choice of shipment size along with the discrete choice of mode, as in Inaba and Wallace (1989); de Jong (2009).

A survey of empirical work in the context of international (particularly European) freight transportation is done in Tavasszy (2006), where several directions of innovation are identified: freight economy linkages, logistics behavioral modeling, and freight trips and networks, and the interaction of freight with passenger travel. Similarly, in Wigan and Southworth (2006) shortcomings of current modeling approaches (e.g., reliance on the limited and sparse data available) and practices (e.g., using some models for another purpose from that for which they were designed) are pointed out. Some of these insights are also well summarized and nuanced in Ben-Akiva et al. (2008), where the authors present some of the most advanced developments in freight transportation modeling.

2.2. Freight attributes in the literature

Attributes used to explain variability in mode choice and the ensuing observed modal shares are surveyed in CUTR (2001), with a focus on (mostly disaggregate) discrete-choice models. Table 2.2 presents a selection of studies that sample the vast space of freight demand modeling, and that we have found particularly useful for our modeling effort. Using aggregate Commodity Transportation Survey (1977) data from the U.S. Census Bureau, Abdelwahab and Sargious (1992) develop a simultaneous discrete-continuous model that models both the choice of transportation mode (the discrete variable) and the choice of shipment size. To be able to use freight attributes not present in the CTS dataset, the authors use an external dataset (the MIT Commodity Attribute Data File) and several models from the literature. The same discrete-continuous approach is taken also in Veras (2002), where the authors model the vehicle choice decision process using a multinomial logit model (MNL) Ben-Akiva and Lerman (1985) on aggregate U.S. Commodity Flow Survey data from 1993. More recently improvements (in terms of predictive accuracy) have been proposed for the discrete-choice framework, e.g., by noting that the logistic regression problem can be expressed a special case of a more general neural-network training problem Nijkamp et al. (1999), namely the single hidden-layer, feed-forward network. The authors of that study compare several specifications of logit models with corresponding neural network models using European inter-regional commodity flow data and obtain superior prediction accuracy using the neural network.

Some studies have been carried out that use stated preference questionnaire data - e.g., in Danielis and Rotaris (1999) the authors attempt to quantify the amount to which analysis would be improved by the availability of more freight data by

surveying relevant literature on this subject - however the large majority of research uses revealed preference survey data. For example, Jiang et al. (1999) estimates a nested multinomial logit model using survey data for France to study the choice between private (own fleet) and public (purchased; rail, road, or combined) transportation based on many operational characteristics available only at the individual shipping manager level. In Garca-Menndez et al. (2004) the authors build a disaggregate conditional logit model using survey data on exporting firms in the Valencia area to study the role of cost, transit time, and frequency of shipments on the demand for freight transportation in four economic sectors. Recent disaggregate studies have sought to incorporate more logistics-related attributes (e.g., proximity to points of access to modes and transport chain modal breakdown) in explaining the choice of both mode and shipment size Windisch et al. (2010). In Holgun-Veras and Thornston (2000), the authors investigate the relationship with trip length of tonnage for different modes, and show the correspondences between commodity-based and vehicle-trip-based analyses. In Johnson and de Jong (2011), time and cost are studied as primary drivers of both shipment size and mode choice.

3. MODELING FREIGHT DEMAND

3.1. Discrete choice analysis of mode shares

The Multinomial Logit (MNL) model (Ben-Akiva and Lerman (1985)) is a workhorse of demand and behavior analysis. The setup assumes a set of N individuals, each individual $n \in \{1, \dots N\}$ facing M choices $m \in \{1, \dots M\}$, and maximizing utility functions of the form $U(\cdot) = V_n(\cdot) + \epsilon_n$, with $V_n(\cdot)$ a deterministic component (usually a linear function of covariates), and $\epsilon_n \sim Gumbel(0,1)$ a random component. The predominant functional form of the deterministic utility V is a linear-in-parameters one with the vector **x** of attributes, $V(x_n) = \boldsymbol{\beta}^T x_n$. In this case it can be shown that the probability (and ensuing market share) of mode m under the model is

$$P_n(m) = \frac{e^{\beta^T x_{nm}}}{\sum_j e^{\beta^T x_{nj}}}, \tag{1}$$

and the relative probability of choosing mode m with respect to mode j is given by the *log-odds ratio*

$$\log \frac{P_n(m)}{P_n(j)} = \beta(x_{nm} - x_{nj}). \tag{2}$$

We chose the MNL specification because of its wide acceptance in the literature, its closed-form expression of the choice probabilities, and its relative ease of estimation.

3.2. Specification of demand utility function

Following the discussion in Section 2.2 we focus on two of the most important factors affecting freight demand: unit cost (perceived on the demand-side of freight transportation) and haul time. Thus the linear-in-parameters specification of the utility function (see above) takes the following form:

$$V_i^m(\cdot) = \beta_0^m + \beta_1 \, \text{VALUE}/_{\text{TON}} \times \text{TIME} + \beta_2 \, \text{COST}/_{\text{TON} - \text{MILE}} + \sum_j \delta_j^m \mathbf{1}\{\text{Origin} = j\} +$$
$$\sum_j \gamma_j^m \mathbf{1}\{\text{Destination} = j\} + \sum_j \eta_j^m \mathbf{1}\{\text{Market} = j\} \tag{3}$$

where i refers to the observation and m to the mode. The variables included (and that we need to estimate as inputs) are:

- TIME^m is the transit time in hours for mode m. It is expected that larger times of haul will decrease utility.

- $\text{VALUE}/\text{TON}^m$ is commodity unit value in \$/ton. Manufactured and high-end commodities (such as electronics or

  pharmaceuticals) have larger unit value, whereas bulk commodities (e.g., coal or grains) have lower unit value (see Section 4). While unit value is a characteristic of the commodity, and not the shipping mode, their product $(\text{VALUE}/\text{TON}^m \times \text{TIME}^m)$ is a rough proxy for the capital value tied-up in transit, which both shippers and

  receivers would like to minimize.
- $\text{COST}/\text{TON} - \text{MILE}^m$ is the unit cost in \$/ton-mile for mode m. We also estimate specifications where unit cost

  is expressed in \$/ton $(\text{COST}/\text{TON}^m)$ to compare with literature estimates that use this formulation.

- $\gamma^m, \delta^m, \eta^m$, are coefficients on dummy variables corresponding to origins and destination regions, and commodities (markets). Using such fixed effects accounts for more variation in the freight flow data that is not captured in either cost or time.

We describe our estimation of the above inputs to the model in Section 4. We test several specifications of the model by choosing subsets of the above variables in 5.

3.3. Model estimation strategy

Our data is in the form of aggregate flows between origin and destination regions by commodity and by mode (see Section 4). Thus the typical Maximum-Likelihood estimation used in discrete choice analysis Ben-Akiva and Lerman (1985) is not directly applicable. Thus we turn to the alternative method of estimation using ordinary least squares (OLS). Berkson's OLS method Ben-Akiva and Lerman (1985) estimates the model as follows:

$$\log \frac{P_n(i)}{P_n(j)} = \beta\left(x_{ni} - x_{nj}\right) + \epsilon_n,$$

with i and j two alternatives from the choice set, x_i and x_j their corresponding attributes, and $P_n(i)$ and $P_n(j)$ the observed shares. Here alternative j is the reference alternative. Note that an efficient estimator needs to account for the struc heteroscedasticity of ϵ_i (see below). The main advantages of OLS estimation of MNL are its reduced computational bu and is thus suited for large datasets. However, as pointed out in Section 4, the basic MNL model does not account for highly-skewed data due to the presence of many zero flows. To account for heteroskedasticity, we used a formulation of Feasible Generalized Least Squares (FGLS) with a flexible, multiplicative structure on the error covariance term due to Harvey Greene (2002): (4)

$$E[\epsilon_i \epsilon_i^T] = \sigma^2 \exp(\mathbf{x}^T \boldsymbol{\gamma})$$

This formulation is estimated using two-stage feasible generalized least squares Greene (2002).

3.4. Model inputs estimation (5)

Following existing literature on freight demand modeling, we set to estimate major determinants (inputs to the utility equation (3) that may be reasonably expected to enter shipping decisions (for air, rail, and road transport), on a per-mode basis:

- **Unit cost.** This represents the cost paid by the beneficiary of the freight transportation service to transport one more unit of a particular good, for one more mile. We estimate marginal unit cost using financial revenue or expenses statements and survey data. We assume perfect markets (zero profits) in which costs reflected to the receivers are equal to the expenditures incurred by the shippers.
- **Haul speed/time.** Modal speed and, related, time-of-haul are critical in shipping, as they affect multiple interrelated decisions (e.g., mode choice and shipment size). While primarily capturing a main characteristic of the mode of transportation, both quantities can also be seen as a measure of the quality and availability of the infrastructure between given origin and destination points.
- **Effective distance.** Distance is a fundamental determinant of freight transportation demand Holgun-Veras and Thornston (2000) which also captures infrastructure availability and quality between given origin-destination pairs for given modes.

Here we develop cost and speed curves as function of distance using operations and financial statements data for various modes. For each mode in our analysis (air, rail, road) we fit a functional form with distance as below:

$$y_i = a + \frac{b}{x_i - c} + \sum_j \delta_j \mathbf{1}\{\text{Origin}_i = j\} + \sum_j \gamma_j \mathbf{1}\{\text{Destination}_i = j\} + \sum_j \eta_j \mathbf{1}\{\text{Market}_i = j\} + \epsilon_i \quad (6)$$

where i refers to the observation, y_i is either unit cost or average haul speed, x_i is distance, and $a, b, c,$ and $\delta_i, \gamma_i, \eta_i$ are coefficients to be estimated. The functional form (6) captures certain key economic phenomena for the distance dependence of unit cost or speed (e.g., Schaefer and Victor (1997); Ben-Akiva et al. (2008)):

- A unit cost that is inversely-proportional with distance captures the effect of economy-of-scales and market size: larger haul distances are typically serviced by large companies, which in turn achieve economies of scale by transporting more goods to more markets.

- In the case of speed, an increasing door-to-door speed with distance for trucking captures the effect of switching from local (low-speed) roads (usually for small urban shipments) to highways for inter-city freight. This phenomenon also applies to railway and air freight transport, since for longer haul distances the maximum admissible speed on particular segments is maintained for a longer period of time, and acceleration and breaking periods (which, in fact, amount to a solid share of the energy consumed) are a relatively smaller share of the total travel time.

Although a strong dependence of both cost and speed with distance is expected, distance alone does not explain much of the geographic and market-specific variability observed in these parameters. Effects such as operational decisions, region-specific infrastructure quality and regulation constraints, or employment and industry output characteristics for particular regions are all absorbed in the respective dummy variables in Equation (6).

4. DATA DESCRIPTION AND INTEGRATION

In this paper we focus on freight transported by air, truck, and rail primarily because these three modes are representative of single modes used in shipping, and amount for most freight tonnage, but also because of data availability constraints. We deliberately leave out one important mode - intermodal shipping, because *(i)* it has lost significant market share over the past 20 years (from 27% in 1990 to 12% in 2010), *(ii)* our focus is on single modes, and *(iii)* it is comparatively difficult to obtain relevant cost and speed data for intermodal freight. We make use of several types of data from a number of publicly-available sources to arrive at a consistent dataset describing U.S.-wide freight transportation demand and key factors determining it. Mainly we are dealing with two types of data: operational (describing financial and logistic decisions) and freight flows (aggregate tonnage and dollar values shipped between regions, broken down by commodity classes and mode of transport). Note that some sources contain disaggregate data (reflecting actual decisions by shippers), whereas others contain aggregate data (totals and averages across many firms or shipments). Below we describe in detail these data and highlight the caveats of using the respective sources in discrete-choice models. All origin and destination information is reported at the level of the FAF[3] analysis zones (see below); we matched all data not in this form from all our sources outlined in this section. All physical values (tonnages, speeds, infrastructure distances) reported are for year 2007, the date for which the latest five-yearly Commodity Flow Survey was available (see below). All monetary values have been converted to 2007 U.S. $ units.

4.1. Freight flow data

Commodity Flow Survey (CFS). This is the primary publicly-available freight flow data resource Bureau (2007) administered by the Census Bureau. It is conducted every 5 years (the latest version available is for 2007). The data is collected via a mandatory questionnaire sent at random to economic establishments in every U.S. state, according to a methodology that ensures national representativeness of the obtained data. The CFS contains flows for a number of modes (air, rail, truck, water, parcels), including origin-destination (at the economic area or state level) and commodity class (up to 5-digit STCG disaggregation) or industry sector information. The data is reported at different levels of aggregation, and at the most disaggregate levels there are many missing data points due to estimation errors or concerns about commercial sensitivity. For that reason we did not use this dataset in our estimation, but resorted to the FAF[3] dataset below.

Freight Analysis Framework 3 (FAF3). This dataset has been prepared by the Transportation research group at the Oak Ridge National Laboratory Oak Ridge National Laboratory (2010) and its stated goal is to address the shortcomings of the CFS. In particular, the authors focus on obtaining estimates for the "missing" data points in the most disaggregated tables in the CFS by using data from higher aggregation levels, Input-Output tables from the Economic Census Bureau (2007), and econometric modeling. The FAF[3] contains aggregate (at the economic area level) tonnage and dollar flows reported for each of 123 origins and destinations, 8 modes (which includes air, rail, truck, water, and parcel) and commodity class (43 classes at the SCTG 2-digit level), i.e., on a $123 \times 123 \times 43 \times 8$ grid. It reports about 32% more tonnage than the CFS, mainly because it contains data for additional modes (e.g., pipelines) and commodity classes (e.g., agricultural products and oil) that are out-of-scope for the CFS. In our analysis in Section 5 we used the freight flows as reported in the FAF[3]. The modal breakdown in the FAF[3] dataset is summarized in Table 2. Road transport (mainly trucks) account for 76% of the total tonnage transported, which shows an economy dominated by truck transportation (in 2007). Rail freight makes up for 13% of tonnage, but only 3% of value, confirming that rail is used for transporting mainly bulk, cheap commodities (e.g., coal or grains). Air freight is less than fraction of a percent of total tonnage, yet accounts for about 1% of value transported, which is consistent with the picture of air transporting high-value commodities such as electronics and pharmaceuticals. Intermodal freight accounts for 3% of total tonnage (but 11% of value); being much smaller than single modes we left it out of our analysis. It is

also instructive to look at a breakdown with distance of the modal split as presented in Figure 1. There we illustrate the cumulative tonnage percentage by each mode as a function of distance (in thousand miles, see distance calculation discussion in Section 4.2 below). The heterogeneity between modes from the haul distance is clear, with road freight reaching more than 95% of tonnage for distances of 500 miles, whereas rail and air freight take, in turn, 1500 and 2500 miles to reach that tonnage percentage.

Even the relatively coarse level of aggregation of the FAF[3] may be too detailed for a high-level analysis of freight trends U.S.-wide. This is noted in several studies in the literature, e.g., Ming (2011) and the references within. We used the aggregation from the 43 two-digit level SCTG classes to 12 commodities as proposed in that study (see Table Appendix A in the Appendix). The basis for aggregation is the similar sectorial input/output patterns exhibited by commodities in each aggregation group. Commodity-based analysis is an important paradigm in freight transportation Ben-Akiva et al. (2008) as the heterogeneity due to sectorial input/output differences, different time sensitivities (e.g., perishable vs. non-perishable commodities) and unit value can be better captured by explicitly including commodity information in the modeling process. In Figure 2 we present a breakdown of the modal shares with distance and by commodity class (as proposed by Ming (2011). Clear patterns emerge of the alternation between dominant modes at different haul distances, with truck usually dominating small distances, rail taking over at medium distances, and air becoming dominant for very large distances. However this picture is an aggregate, and therefore hides important differences at particular commodity classes. For example, coal and fuels are primarily transported by rail for up to medium distances, by truck for larger distances, and then only trace quantities are transported by air (e.g., from the lower U.S. states to Alaska) for very large distances. However higher-valued goods such as tobacco are transported predominantly by road up to very long distances, when air transportation starts to become a sizable alternative.

Mode	VAL [$M]	VALP [%]	TON [000's]	TONP [%]
Truck	10,223,022	77	12,580,080	76
Rail	373,963	3	1,745,364	11
Water	98,221	1	356,658	2
Air (include truck-air)	152,229	1	2,679	0
Multiple modes & mail	1,681,048	13	522,360	3
Pipeline	552,231	4	1,100,224	7
Other and unknown	257,486	2	268,798	2

Table 2: Tonnage and value distribution across modes for the FAF[3] dataset.

4.2. O-D distance calculations

As of version 3:3 (released in July 2012), the Freight Analysis Framework dataset Oak Ridge National Laboratory (2010) also includes estimated ton-mileage for origin-destination-commodity-mode (123 123 12 3) tuples that contain non-zero freight flows. For these tuples we estimated the average distance travelled by freight for a given mode as

$$\text{Effective Distance} = \frac{\text{Ton} - \text{Miles}}{\text{Tons}} \qquad (7)$$

For validation we compared these estimates with estimates using a network model of the road and rail infrastructure impedances Southworth and Peterson (2000). As the network model reported distances at the U.S. county level, we matched the counties to FAF[3] zones using county administrative and geographic information from the U.S. Census Bureau. To estimate air distances we used average great circle distance calculations between airports within FAF[3] zones. We split our distance calculations into those OD-pairs that do not include legs in Alaska, and those that do. This was motivated because of the particular geographical situation of Alaska, which is separated from the other 48 contiguous states in the U.S. We compare the two estimates in Figure 3. Note the good agreement between the two methods (correlation coefficient > 0:9 for all modes). Thus we used the network model estimates of average distances for those FAF[3] regions for which no freight flows were reported (structural zeros).
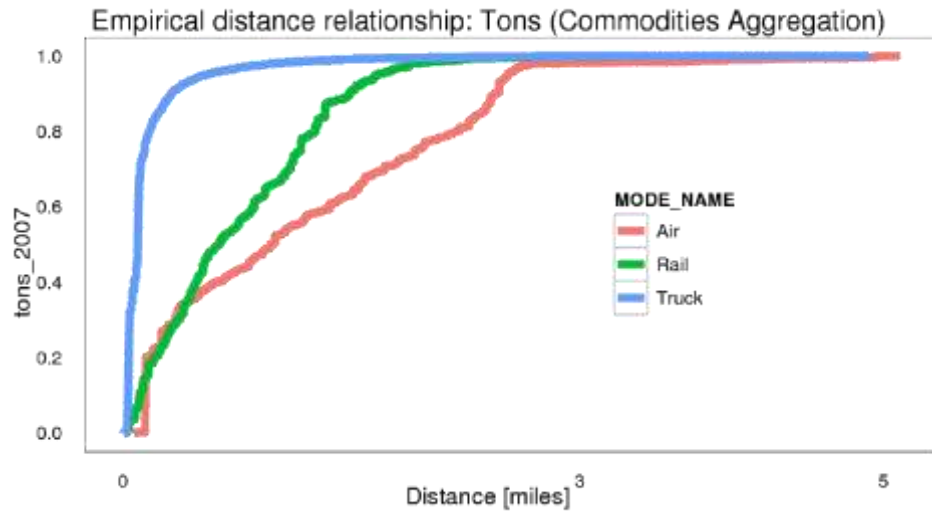
Figure 1: Cumulative tonnage breakdown by mode as function of distance (thousand miles).
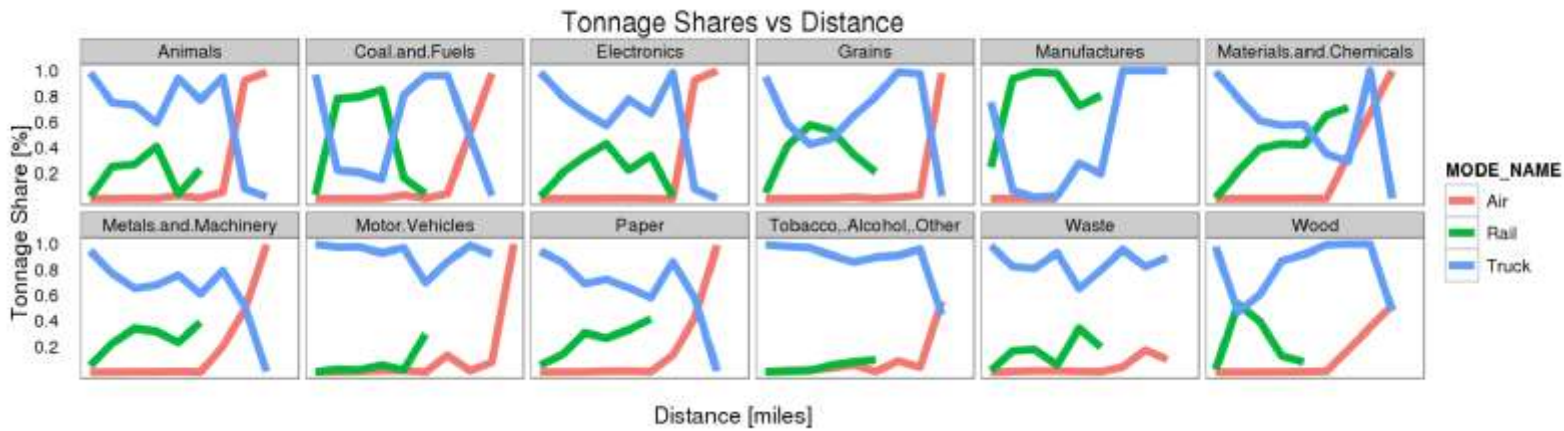

Figure 2: Commodity-based analysis: modal shares dependence with distance (0-5000 miles).
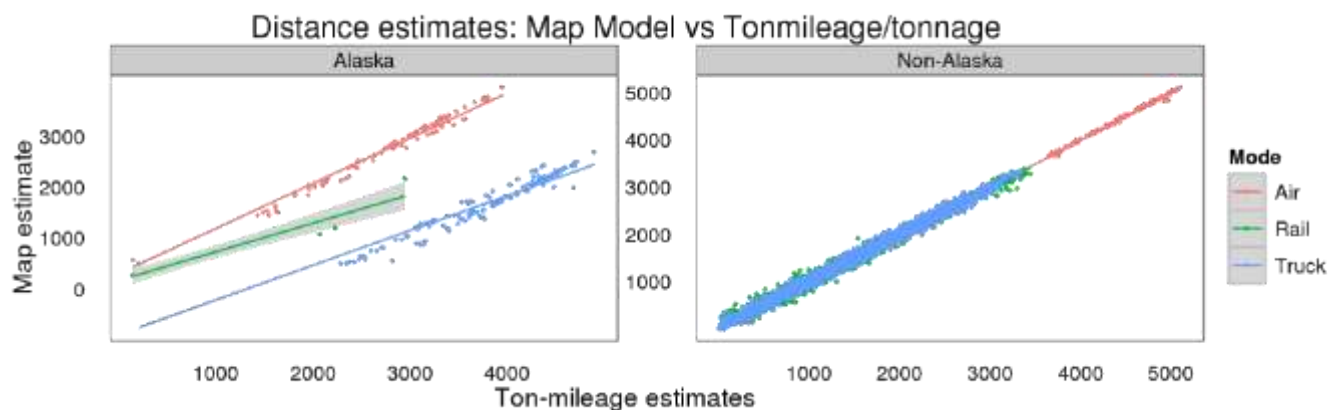
Figure 3: Distance calculations using a map model Southworth and Peterson (2000) and comparison with effective distance derived from ton-mileage figures Oak Ridge National Laboratory (2010).
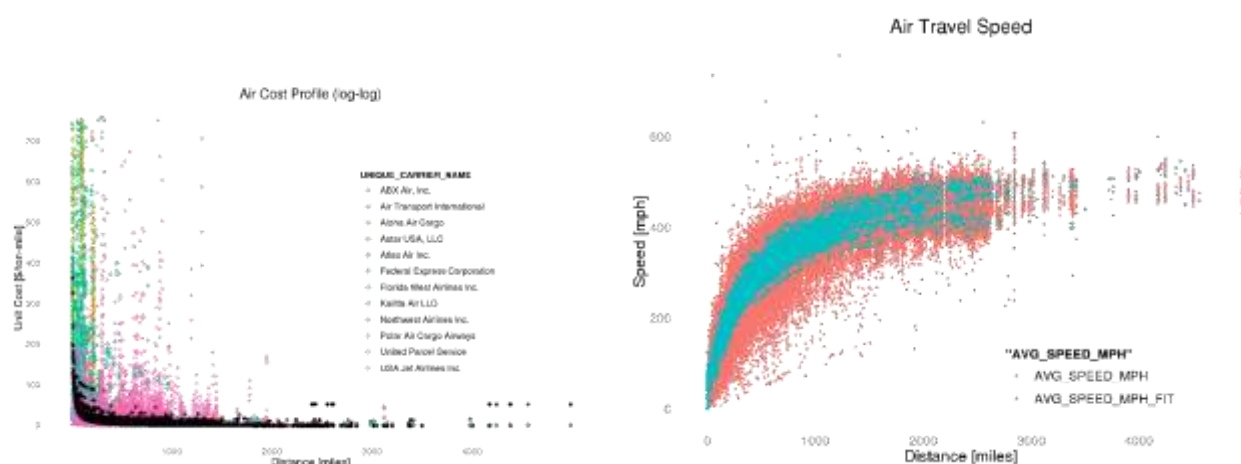

Figure 4: Air freight attribute estimates: unit cost (left) and speed (right).

4.3. Air freight attributes

Unit cost. We use the Air Transportation Authority Form 41 (Schedule P-1.2) for 2007 RITA (2011) available online at the Research and Innovative Technology Administration (RITA) website [1]. This database contains quarterly financial statements for large operators (those with operating revenues of at least $20 million), including operating revenue and expenses, income tax, depreciation and amortization costs etc. We use carrier operating revenue as a proxy for the costs incurred on the demand side for transporting freight. Aggregate revenue data (over each quarter of 2007) are reported for around 30 major carriers that transport freight, for several types of aircraft the carriers operate. We then matched the revenues data with data on freight tonnage transported as reported in Schedule B2.1 of RITA (2011) by airline and airframe type. The calculations are summarized in Figure 4 (left panel). While there is much variance that is not explained by distance, the general declining trend with distance is clearly observable. According to the discussion in Section 2.2, we fit a functional form similar to Equation 6 to the unit cost data (*t-test* results are summarized in Table A.5, left):

(8)

$$\text{cost}_i = a + \frac{b}{\text{distance}_i} + \sum_j \delta_j \mathbf{1}\{\text{Origin}_i = j\} + \epsilon_i$$

Haul speed. The Air Transportation Authority Form 41 (Schedule T-100) RITA (2011) contains quarterly operations data for airlines operating in the U.S. We have used the aggregated (quarterly totals) data tables for domestic carriers containing details on non-stop flight routes (reported on a per-airport basis), including carrier, U.S. origin and destination airport, amount of freight and number of passengers, number of departures performed, number of ramp-to-ramp hours. We computed average haul speed for those segments and air carriers that carried freight (excluding passenger flights). For the purpose of regression analysis we matched each airport to its corresponding FAF[3] zone. The results are summarized in Figure 4 (right panel), where each point represents an O-D segment with nonzero freight activity. As before, we t a functional form similar to Equation 6 to the aircraft speed data (results are summarized in Table A.5, left):

$$\text{speed}_i = a + \frac{b}{\text{distance}_i - c} + \sum_j \delta_j \mathbf{1}\{\text{Origin}_i = j\} + \sum_j \gamma_j \mathbf{1}\{\text{Destination}_i = j\} + \epsilon_i \qquad (9)$$

4.4. Rail freight attributes

Unit cost. The most comprehensive public dataset that details railway operations cost is the Surface Transportation Board Carload Waybill Survey RAILINC (2010), the newest version of which is for 2007.
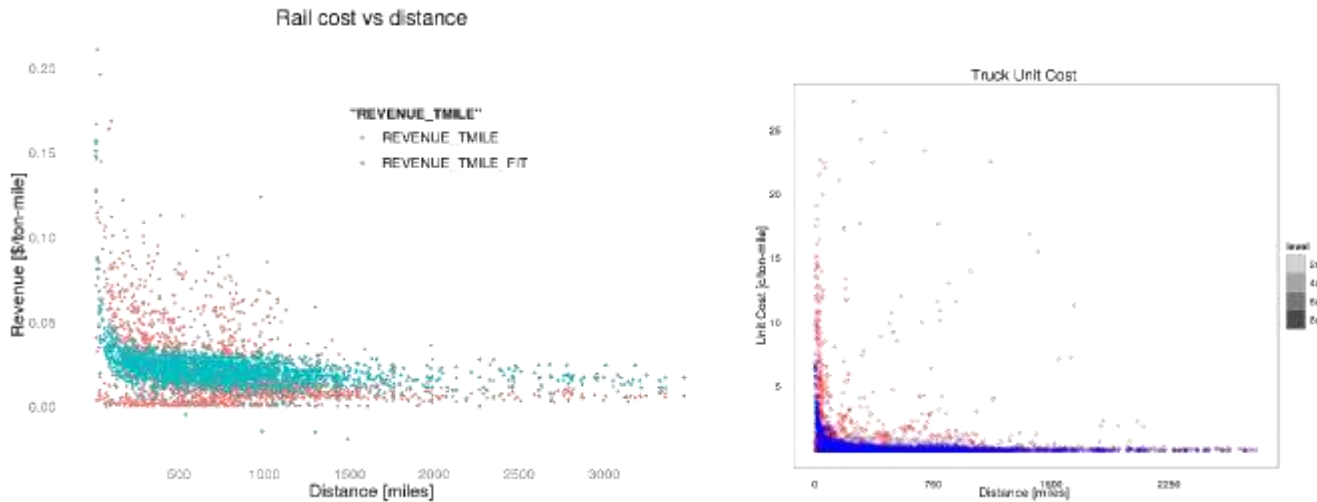

Figure 5: Unit marginal cost functional form estimate. *Left*: Rail unit cost data estimated from the 2007 Carload Waybill Survey (public sample); *Right*: Truck unit cost data from the Motor Carrier Annual Report (see text).

All major (operating more than 4; 500 carload annually) U.S. rail carriers are required to submit answers to the survey, which makes this dataset the most comprehensive public source on the topic. However only a subsample of this dataset is available on line because it contains commercially-sensitive information; to use the complete data special permission must be obtained. We have only used the publicly-available data. The survey contains rich disaggregate information on shipments, including length of haul, revenue from shipment, shipment origin and destination, commodity class transported, number of legs in shipment etc. We estimated the unit cost ($/ton-mile) across the shipments in the sample and present the calculations in Figure 5. A summary of the t is given in Table A.5 in the Appendix.

Haul speed. Literature reports very little variation in operational characteristics of rail engines**,** e.g., Gucwa and Schafer (2011). As such we used the verge speed value of 19.3 mph for all the rail freight activity in the U.S. (in 2007) available from the American Association of Railroads Railroad Facts 2007 RAILINC (2008)

4.5. Road freight attributes

Unit cost. To estimate unit cost for road transportation we used the TranStats Motor Carrier Annual Report database TranStats (2007). This dataset is the result of a mandatory survey filled by large (with more than $ 3 million in revenues) motorized operators that contains information on operating revenue and expenditure, fuel used, type of operation, type of goods carried, miles traveled. The most recent data reported is for 2003; we thus applied the appropriate inflation factor (to 2007 values) to all dollar figures. We used schedules S200 (operating income) and S300 (operational statistics) to estimate unit cost (in $/ton-mile) from information on total revenues, ton-mileage, tonnage, and number of shipments. For each carrier we estimated typical freight haul distance from the ton-mileage and tonnage information reported. The results are presented in Figure 6 (left panel). Again we observe the declining unit cost with distance, and fit the appropriate functional form (6)

$$\text{cost}_i = a + \frac{b}{\text{distance}_i} + \sum_j \delta_j \mathbf{1}\{\text{Origin}_i = j\} + \epsilon_i \tag{10}$$

and report the obtained values for a and b in Table A.6 (right panel).
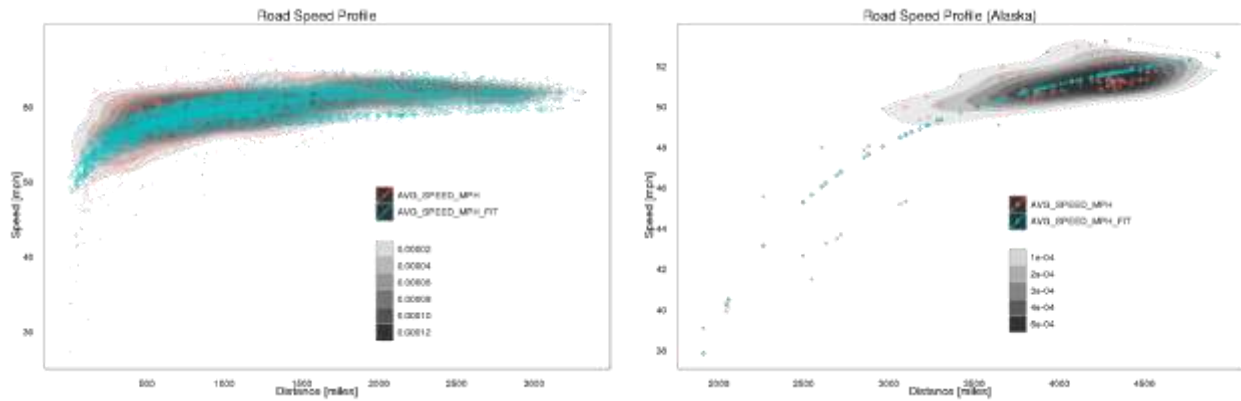

Figure 6: Road freight transportation speed estimates (left: U.S. without legs in Alaska; right: Alaska).

Haul speed. To understand the distance dependence of road transportation, we used road segment typical speed values obtained from a state-of-the-art online map model API (Google Maps). We queried the API scriptically for road travel distances for several cities within and between the 123 FAF[3] zones, in a similar fashion as Wang and Xu (2011). The resulting speed profiles are illustrated in Figure 6, where each red dot corresponds to an origin-destination FAF3 zone excluding zones in Alaska (middle panel) and only including Alaska pairs (left). As expected, the average speed increases with distance, and levels off at about 60 mph, which is the legal limit on most highways in the U.S. For the two cases (Alaska, non-Alaska) we fit the familiar functional form with distance

$$\text{speed}_i = a + \frac{b}{\text{distance}_i - c} + \sum_j \delta_j \mathbf{1}\{\text{Origin}_i = j\} + \epsilon_i \tag{11}$$

and summarize the results in Table A.6.

5. MODELING MODAL MARKET SHARES

In this section we present an application of the dataset assembled as described in Section 4 to estimating a discrete-choice model (multinomial logit) as introduced above in Section 3. We calculated typical quantities of interest in the freight transportation context (attribute coefficients, elasticities, value-of-time) both overall and on a per-commodity basis. We compared results obtained using several estimation methodologies (Ordinary Least-Squares and Feasible Generalized Least-Squares).

5.1. Caveats of the data and the model

The economic forces giving rise to the demand for freight transportation in the U.S. are highly complex, and we have only crudely modeled them here. For example, the model formulation as in (4) does not account for the fact that many flows on certain modes and commodities do not exist between all origin-destination pairs (i.e., are structurally zero). As an example, clearly there is no coal being transported by air from Texas to Hawaii. This phenomenon is clearly not only related to the combination of attributes (unit cost, speed, distance, unit value) as described above in Section 4); it rather arises because of geographical and infrastructure constraints, exogenous market forces, and specific sectoral input-output relationships interacting with spatially-distributed supply and demand, for instance. Granular data that allow identification of such effects requires a considerable effort to obtain and integrate. The structural zeros issue may be alleviated by either i) explicitly modeling the absence of flows (e.g., by correlating it with sectoral input/output, employment, and population patterns over the U.S. geography), or ii) by aggregation of finer-grained O/D flows into larger, coarser geographical regions (to reduce variance and increase homogeneity in economic conditions). As our emphasis was not on developing a detailed freight demand model, here we have adopted the latter approach. The effect of aggregation is visible in Figure 7, where the (log-scale) distributions of model inputs are illustrated. Note that for the 9×9 geographical region O/D pairs (right panel) the distributions are much less skewed than for the 123×123 FAF[3] (left panel).

Variable	Estimate OLS	P.Value OLS	Estimate GLS	P.Value GLS
COST TMILE	-2.15 (0.306)	<0.001	-1.97 (0.327)	<0.001
DISTANCE	-0.00405 (0.00206)	0.050	-0.00393 (0.00196)	0.050
MODE NAME Air	-10.9 (0.498)	<0.001	-10.8 (0.495)	<0.001
MODE NAME Rail	-4.85 (0.455)	<0.001	-4.8 (0.428)	<0.001
Region DEST Middle Atlantic	1.49 (0.469)	<0.001	0.88 (0.474)	0.060
Region DEST New England	-1.19 (0.43)	0.010	-1.26 (0.465)	0.010
Region DEST Pacific	3.27 (0.479)	<0.001	3.17 (0.469)	<0.001
Region DEST West South Central	0.808 (0.431)	0.060	0.832 (0.391)	0.030
TIME HRS	-0.0201 (0.00652)	<0.001	-0.0221 (0.00657)	<0.001
TIME HRS :VALUE TON	-0.000205 (2.91e-05)	<0.001	-0.000199 (2.39e-05)	<0.001

Table 3: Log-odds estimation of MNL (Cost in Ton-Miles)

5.2. Model estimation results

We first estimated the discrete-choice model with the utility specification (3) using Ordinary Least Squares, and report the results in Table 3. The reference alternative in Equation (4) necessary for identification of estimates was trucking. As suggested by the OLS residuals dependence with left-hand-side covariates illustrated in Figure 8, there is heteroskedasticity in the data in particular on the unit value and unit cost variables, which is also indicated by a Breusch-Pagan test Greene (2002) (95% confidence level). To alleviate this issue and increase estimation efficiency, we recomputed both the coefficient estimates and the standard errors using a Feasible Generalized Least Squares approach. These estimates are also shown in Table 3. All estimates were done on a 9×9 O/D dataset (Figure 7, left panel). All the coefficients have the expected signs (negative for unit cost, distance, time, and capital tied up in transit). The negative coefficients on the alternative-specific constants for air and rail are relatively large and negative, which indicates that a good part of the heterogeneity in the data is not explained by the studied freight attributes alone. Moreover, controlling for all factors, air and rail are less desirable than

truck, as shown by the relatively large alternative-specific constants on those modes. This result is intuitive given the door-to-door operation characteristics inherent to road transportation. Note that coefficient estimates are statistically significant at least at the 0.01 level.

VOT Unit	OLS	GLS	Aggregation
($/ton-mile-hour)	0.009	0.010	9 x 9
($/ton-hour)	12.400	13.600	9 x 9
VOT Unit	Value	Study	Country
($/ton-hour)	$6.56 – $10.59 (depending on commodity)	de Jong (2004)	Netherlands
($/ton-hour)	$1 – $110 (depending on commodity)	de Jong (2004)	Sweden
($/ton-hour)	$41 – $53 (depending on commodity)	Bergkvist (2001)	U.K.

Table 4: Value of time estimates

A comparison with literature estimates is not that immediate because of the large heterogeneity in data type and model specification employed in various studies. We found relatively little work on national-level econometric studies that investigate the effects of unit marginal cost and measures of infrastructure availability such as haul time over the entire U.S. geography. Our interest here was to compare signs and relative magnitudes within-studies; cross-study analysis is not straightforward because of the wide variation in model setup and data: different types (aggregate, disaggregate), countries (U.S., Europe, Latin America, etc.), units system (metric, imperial), different currencies (U.S., SEK etc.) and so on. In general, much of the heterogeneity is not explained by used attributes alone, as indicated by the large relative magnitudes of the alternative-specific constants. This observation is consistent with our own modeling results. A significant number of the studies we surveyed did not report any value for the ratio of explained variance R^2, but for those that did this quantity varied greatly in magnitude. Note that the quantity depends on the estimation methodology - typically the OLS R^2 values will be higher (0:6 0:8) than the pseudo-R^2 (or McFadden's Ben-Akiva and Lerman (1985)), which typically are on the order of 0:2 0:4 Ben-Akiva et al. (2008). Our result (R^2 = 0:79) is thus typical of these kind of models and specifications.

Table 4 presents calculations of the value-of-time (VOT) of freight transportation for the U.S. in 2007. We estimate value of time as

$$\text{VOT} \left[\frac{\$}{\text{ton} - \text{mile} \times \text{hour}} \right] = \frac{\beta_{\text{TIME}}}{\beta_{\text{COST}}} \tag{12}$$

where COST and TIME are the regression coefficients for unit marginal cost and typical time, respectively.

The VOTs obtained via OLS and FGLS were of similar magnitude. Since some studies in the literature reported VOT in $/ton-hour, we repeated the model estimation with unit cost expressed in $/ton instead of $/ton-mile (not shown here for space reasons). A back-of-the-envelope calculation multiplying the result expressed in $/ton-mile-hour with the average tonnage flow reveals a similar result. Our result (using both OLS and FGLS) of about $ 13/ton-hour falls within the range of values reported in the literature for other countries, as presented in Table 4.

6. DISCUSSION

The main goal with this exercise was to illustrate the integration and use of public data in per-forming analysis relevant to transportation and energy policy at a U.S.-level, and to identify certain points where the availability of additional data could make a positive difference. We surveyed a large body of literature on econometric analyses of freight transportation in different countries and regions, using different types of data, etc. We first presented our work on integrating many publicly-

available datasets related to freight transportation in the U.S. We arrived at a database containing information on key attributes of demand for freight - cost, time (speed), and distances - that we used to estimate a simple discrete-choice model of modal shares. We then calculated the Value-of-Time for freight ($13/ton-hour) and found that it is broadly in agreement with values reported in the literature. However that both reflects the appropriateness of this approach as well as the large variance of estimates across different markets and geographies present in the literature. Based on our experience with the process, we can readily identify several shortcomings:

Data availability. Collecting transportation data, and freight data in particular, is expensive, as identified by other previous studies Ben-Akiva and Lerman (1985); Ben-Akiva et al. (2008). Moreover, freight companies and professional associations that we contacted to request specific data on logistics and financial operations for our research were generally reluctant to provide that information. This is quite understandable, since that information usually has high value for competitive advantage and commercial reasons. Absent granular information at the level of individual shipments, one needs to rely on aggregates that mask away important trends. This is not the case for all industries though; for example there is much richer air transportation data available publicly than surface transportation data (trucking or rail). One reason behind this is that air travel is more severely regulated (for safety reasons) than other modes; moreover trends in that industry are that some type of freight (e.g., mail or packages) is generally transported alongside passengers on commercial flights, for which airlines are required to report a wealth of data, including departure and arrival times, tonnage carried, delays, fuel used etc. Working with major industry players in other sectors of transportation (railway and motor carrier operators) to make more data available for research would spur much interest and innovation, as it is the case with recent initiative in the air transportation industry (e.g., the GE flight challenge[5]). This is particularly the case in light of the recent increasing trends of e-commerce and the afferent business it produces for freight operators who need to manage carefully and flexibly their operations.

Data heterogeneity and integration. As expected, the data we used in our estimations varied greatly in scope, method of acquisition, quantity, granularity, etc. The freight flow data (the FAF[3] database Oak Ridge National Laboratory (2010)) was itself the product of an extensive modeling e ort, pulling together data from a variety of sources. Data on costs and operations was either disaggregate in nature (e.g., individual cargo loads for the Carload Waybill Survey RAILINC (2010)) or aggregate over the whole country (e.g., the average speed figure for rail RAILINC (2008)). To circumvent in part the large degree of heterogeneity in the different inputs to our discrete-choice analysis, we adopted the model (6) of an inverse-relationship with distance of the studied quantity as described in Section 3. As it is clear from the discussion there, this relationship captures a general trend, but even with the addition of fixed effects in the regression models much of the variance in the data remains unexplained. One simple way of dealing with the heterogeneity that we adopted was aggregation (see discussion in Section 5 and Figures 7 in the Appendix) was aggregation. However, the better approach is to attempt to model this variability explicitly in terms of additional covariates on the state of the economy or infrastructure. Clearly that would require additional data collection and duration efforts.

In conclusion, we argue for more research in the creation and maintenance of standard databases that allow more transparency of analysis, by integrating granular data from multiple sources related to freight transportation. We believe that such tools will move the field of freight transportation econometrics into a possibly transformational phase, as it has been the case for other fields of inquiry that have experienced increased data availability over the last years.

REFERENCES

Abdelwahab, W. and M. Sargious (1992). Modelling the demand for freight transport: A new ap-proach. Journal of Transport Economics and Policy 26, 4970.

Authority, E. I. (2011, 6). Energy data book. Available at: http://www.eia.gov/. Last accessed June 2011.

Ben-Akiva, M. and S. R. Lerman (1985). Discrete choice analysis. MIT Press.

Ben-Akiva, M., H. Meersman, and E. Voorde (2008). Recent developments in transport modelling: lessons for the freight sector. Elsevier.

Bergkvist, E. (2001, August). The value of time and forecasting of owsin freight transportation. ERSA conference papers ersa01p271, European Regional Science Association.

Bureau, U. C. (2007). Commodity ow survey. Available at: http://www.census.gov. Last accessed June 2011.

Clark, C., H. T. Naughton, B. Proulx, and P. Thoma. A survey of the freight transportation demand literature and a comparison of elasticity estimates.

CUTR (2001). Analysis of freight movement mode choice factors. report for orida department of transportation rail planning and safety. Technical report, The Center for Urban Transportation Research at the University of South Florida.

Danielis, R. and L. Rotaris (1999). An analysis of freigtht transport demand using stated preference data: a survey and a research project for the friuli-venezia giulia region. Transporti Europei (13).

de Jong, G. (2004). Main survey into the value of time in freight transport by road. Prepared for the avv transport research centre, RAND Corporation.

de Jong, G. (2009). Discrete mode and discrete or continuous shipment size choice in freight transport in sweden. ETC Proceedings.

F.M., V. and M. E.K. (2000). Improving the energy e ciency of freight in the united states through commodity-based analysis: justi cation and implementation. Transportation Research Part D: Transport and Environment 5 (1), 11{29.

Garca-Menndez, L., I. Martnez-Zarzoso, and D. P. D. Miguel (2004, September). Determinants of mode choice between road and shipping for freight transport: Evidence for four spanish exporting sectors. Journal of Transport Economics and Policy 38 (3), 447{466.

Greene, W. H. (2002, August). Econometric Analysis. Prentice Hall.

Gucwa, M. and A. Sch•afer (2011). Energy intensity in freight transportation - what lies beneath.

Under review.

Holgun-Veras, J. and E. Thornston. An Investigation of the Relationships Between the Trip Length Distributions in Commodity-based and Trip-based Freight Demand Modeling. Number No.1707.

Inaba, F. S. and N. E. Wallace (1989, November). Spatial price competition and the demand for freight transportation. The Review of Economics and Statistics 71 (4), 614{25.

Jiang, F., P. Johnson, and C. Calzada (1999). Freight demand characteristics and mode choice: An analysis of the results of modeling with disaggregate revealed preference data. Journal of Transportation and Statistics 2 (2).

Johnson, D. and G. de Jong (2011). Heterogeneous response to transport cost and time and model speci cation in freight mode and shipment size choice. International Choice Modelling Conference.

Levinson, D. M., M. Corbett, and M. Hashami (2005). Operating costs for trucks. IEEE Trans. Visu-alization & Comp. Graphics (Proc. InfoVis). Available at http://ssrn.com/abstract=1736159.

Massiani, J. (2003, August). Bene ts of travel time savings for freight transportation : beyond the costs. ERSA conference papers ersa03p388, European Regional Science Association.

Ming, R. L. (2011). Development of a freight generation model through linear regression: An appli-cation to california.

Nijkamp, P., A. Reggiani, and W. F. Tsang (1999). Comparative modelling of interregional transport ows : applications to multimodal european freight transport. Serie Research Memoranda 0002, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Oak Ridge National Laboratory (2010). The freight analysis framework 3: Overview of the faf[3] national freight ow tables. Batelle, Inc.

RAILINC (2008, June). Railroad facts. Technical report, Association of American Railroads. Available at: http://www.aar.org/StatisticsAndPublications.aspx. Last accessed June 2011.

RAILINC (2010). Reference guide for the 2010 surface transportation board carload waybill sample. Available at: http://www.stb.dot.gov/stb/industry/econ_waybill.html.

Reitser, D. (1977). The total energy cost of freight transport. Energy 3, 493{498.

RITA (2011, June). Transstats air transportation data. Available at: http://www.transtats.bts. gov. Last accessed June 2011.

Samimi, A., K. Mohammadian, and K. Kawamura (2010). Behavioral freight movement modeling.

The 12th International Conference on Travel Behaviour Research.

Schaefer, A. and D. G. Victor (1997). The past and future of global mobility. Scienti c American, 56{59.

Schewel, L. and L. Schipper (2011a). Fossil freight: how much fossil fuel does it take to move fossil fuel? Manuscript in preparation.

Schewel, L. and L. Schipper (2011b). Shop 'till we drop: historical and policy analysis of driving-for-shopping and freight energy use in the u.s. Manuscript in preparation.

Schipper, L., C. Saenger, and A. Sudardshan (2011). Transport and carbon emissions in the united states: The long view. Energies 4 (4), 563{581.

Sorratini, J. A. and J. Robert L. Smith (2000). Development of a statewide truck trip forecasting model based on commodity ows and input-output coe cients. Transportation Research Record: Journal of the Transportation Research Board 1707, 49{55.

Southworth, F. and B. Peterson (2000). Intermodal and international freight network modeling.

Transportation Research Part C: Emerging Technologies 8 (1), 147{166.

Tavasszy, L. (2006, September). Freight modelling-an overvierw of international experiences. TRB Conference on Freight Demand Modelling.

Tavasszy, L., I. Davydenko, and K. Ruijgrok (2009, August). The extended generalized cost concept and its application in freight transport and general equilibrium modeling. Integration of Spatial Computable General Equilibrium and Transport Modelling, Bilateral Joint Seminar.

TranStats (2007). Motor carrier operators: Annual report. Available at: http: //www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=775&DB_Short_Name=Motor% 20Carrier%20Operators. Last accessed June 2012.

Veras, J. H. (2002). Revealed preference analysis of commercial vehicle choice process. Journal of transportation engineering.

Vilain, P., L. N. Liu, and D. Aimen (1999). Estimation of commodity in ows to a substate region: An input-output based approach. Transportation Research Record: Journal of the Transportation Research Board 1653, 17{26.

Wang, F. and Y. Xu (2011). Estimating o-d travel time matrix by google maps api: implementation, advantages, and implications. Annals of GIS 17 (4), 199{209.

Wigan, M. R. and F. Southworth (2006). Whats wrong with freight models? and what should we do about it? Paper presented at the 85th Transportation Research Board Annual Meeting.

Windisch, E., G. C. de Jong, R. van Nes, and S. P. Hoogendoorn (2010). A disaggregate freight trans-port model of transport chain and shipment size choice. European transport modelling conference.

Winston, C. (1981). A disaggregate model of the demand for intercity freight transportation. Econo-metrica 49 (4), pp. 981{1006.

Winston, C. (1982). The demand for freight transportation: models and applications. Transportation Research Part A 17A(6), pp. 419{427.

Winston, C. (1983). The demand for freight transportation: models and applications. Transportation Research Part A: General 17 (6), 419 { 427.

Zlatoper, T. J. and Z. Austrian (1989). Freight transportation demand: A survey of recent econometric studies. Transportation 16, 27{46. 10.1007/BF00223045.

Appendix A. Additional tables and figures

	Variable	Air Speed	Air Cost
1	a***	529.12	-1.7492
		(0.49)	(0.1248)
2	b***	-218481.22	2195.3269
		(901.98)	(16.5529)
3	c***	-462.91	
		(1.68)	
4	Adj.R2	0.92	0.22

	Variable	Rail Cost	Rail Speed
1	a***	6e-04	19.6000
		(0.0024)	
2	b***	1.4052	
		(0.0681)	
3	c***		
4	Adj.R2	0.20	

Table A.5: Right: Air model estimation; Left: Rail model estimation.

	Variable	Truck Speed (non-AK)	Truck Speed (AK)	Truck Cost
1	a***	63.5972	56.9069	0.0185
		(0.0224)	(0.3276)	(0.0069)
2	b***	-5279.4518	-17461.1557	62.2338
		(52.5421)	(1366.9314)	(1.2886)
3	c***	-393.7559	993.3408	-6.2121
		(5.1125)	(68.048)	(0.3423)
4	Adj.R2	0.64	0.86	0.11

Table A.6: Truck Model Estimation.

| Class	SCTG Commodities Included (SCTG code)	Class Description
A	Animals and Fish (1); Animal Feed (4); Meat/Seafood (5)	Animals
B	Cereal Grains (2); Other Agr. Products (3)	Grains
C	Milled Grain Products (6); Other Foodstu s (7); Alcoholic Beverages (8); Tobacco Products (9)	Tobacco, Alcohol, Other
D	Building Stone (10); Natural Sands (11); Gravel (12); Nonmetallic Minerals (13); Metallic Ores (14); Basic Chemicals (20); Pharmaceuticals (21); Fertilizers (22); Chemical Products (23); Plastic/Rubber (24); Nonmetallic Mineral Products (31)	Materials and Chemicals
E	Coal (15); Crude Petrolleum (16); Gasoline (17); Fuel Oils (18); Coal -n.e.c. (19)	Coal and Fuel
F	Logs (26); Wood Products (26)	Wood
G	Newspaper/Print (27); Paper Articles (28); Printed Products (29)	Paper
H	Textiles/Leather (30); Furniture (39); Misc. Mfg. Prods. (40); Mixed Freight (43)	Manufactures
I	Base Metals (32); Articles-Base Metals (33); Machinery (34)	Metals and Machinery
J	Electronics (35); Precision Instruments (38)	Electronics
K	Motorized Vehicles (36); Transport Equipment (37)	Motor Vehicles
L	Waste/Scrap (41)	Waste
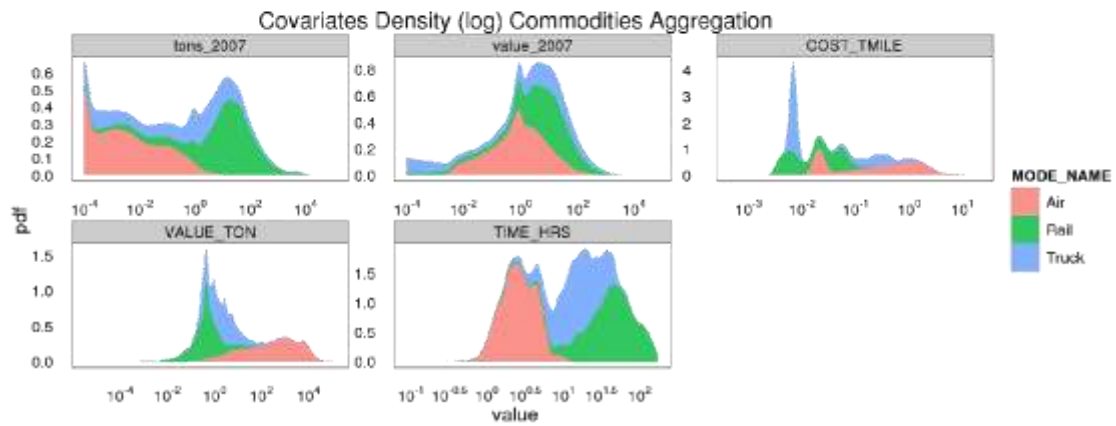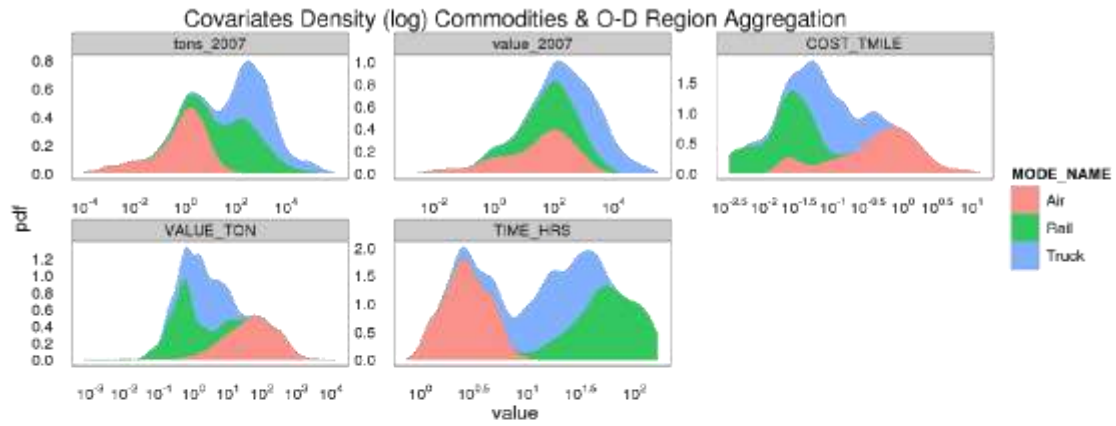
Figure 7: Input freight attributes (log-densities). *Left*: 123×123 FAF[3] Origin-Destination pairs; *Right*: 9×9 Origin-Destination U.S. geographical regions.
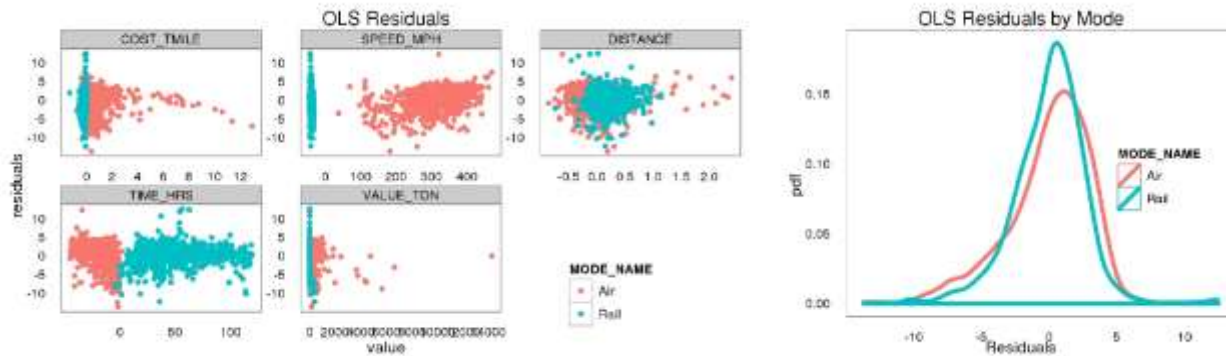

Figure 8: OLS analysis. *Left*: OLS residuals vs Inputs for air and rail. Heteroskedasticity arises primarily because of non-uniform variance in unit cost and speed. *Right:* OLS residuals after heteroskedasticity correction.