Development of Small-Area Origin-Destination Freight Flows for Metropolitan Planning

Submitted for presentation at the Transportation Research Board SHRP 2 Symposium: Innovations in Freight Demand Modeling and Data.

Author: Paul Metaxatos, University of Illinois at Chicago, pavlos@uic.edu

Introduction

Federal legislation has encouraged the consideration of freight movements in state and metropolitan planning processes. Moreover, state departments of transportation (DOTs), metropolitan planning organizations (MPOs), local governments and business leaders have become increasingly aware of the importance of freight transportation for economic development. However, the effective management of freight transportation systems requires sound transportation planning practices, which, in turn, depend on reliable information on local, regional, and national freight distribution that is not readily available today at a desirable geographic scale. This is because freight transportation planning, especially at the metropolitan level, needs flow data at a small geography, say, at the traffic analysis zone (TAZ) or the county level at a minimum. Yet, existing flow data are only available at a coarse geographic level (e.g., state level), and more detailed data (e.g., county level) are proprietary with unknown quality and certainly not suitable for freight route planning. Moreover, MPOs do not have the resources to undertake expensive freight data collection efforts. There is a need, therefore, to develop freight flow data for small areas, that is, at the county level at a minimum or, preferably, at the TAZ level.

This paper presents a method to develop small-area freight origin-destination flows that are suitable for transportation planning using publicly available data. The method draws from the literature in spatial interaction modeling, especially the family of gravity models, and small-area estimation. Small area estimation techniques accommodate single-area or non-flow data. On the other hand, freight movements have, at a minimum, an origin and a destination and spatial interaction models are ideally suited to handle bi-directional flows. The discussion below will show that these two statistical paradigms working together provide a sound theoretical framework for developing (synthetic) freight data at a desirable geographic level.

Spatial Interaction Modeling of Freight Flows

A modern version of the gravity model is $N_{ij} = T_{ij} + \varepsilon_{ij}, \forall i, j$ where $N_{ij}$ is the flow from origin $i$ to destination $j$, $\varepsilon_{ij}$ an error term, and $T_{ij} = E(N_{ij}), \forall i, j$, the expectation of $N_{ij}$, is written as $T_{ij} = A_i B_j F_{ij}, \forall i, j$. In the last expression, $A_i$ and $B_j$ reflect activity at $i$ and $j$, respectively, and $F_{ij}$ reflects the difficulty of getting from $i$ to $j$ or of shipping from $i$ to $j$. In the theory of spatial interaction modeling $A_i$’s and $B_j$’s are known as origin and destination functions the specification of which allows for a rich framework for exploratory analysis in numerous applications involving flows of people, goods, money, etc. (Sen and Simith, 1995). For convenience in estimation, $F_{ij}$ is written, without loss of generality, as $F_{ij} = \sum_{k=1}^{K} \theta_k c_{ij}^k$ where, $\theta_k$ are parameters to be estimated and $c_{ij}^k$ are variables measuring separation between $i$ and $j$.

The gravity model has been intensively studied and has been given a sound theoretical basis. In Sen and Smith (1995), the gravity model has been given a sound theoretical basis on the basis of a very small number of intuitively appealing axioms. Note also that if any of the $\theta$’s above goes to $-\infty$, while other terms stay the same, $T_{ij}$ becomes a linear programming solution (Evans, 1973; Senior and Wilson, 1974; Erlander and Stewart, 1990; Sen and Smith, 1995). This fact also yields another theoretical basis for the gravity model.

There is a considerable body of literature discussing the use of gravity models for freight flow estimation (Sen and Pruthi, 1983; Ashtakala and Murthy, 1993; Smadi and Maze, 1996; Black, 1999; Cheu et al., 2003; Metaxatos, 2004; Matsumoto, 2007). Note that extending to freight a model originally justified for passenger
flows also creates an estimation problem. Maximum Likelihood (ML) procedures for estimation are based on the Poisson or multinomial distribution. It is unlikely that freight flows, which combine flows of coal (which move in train-loads) and much smaller deliveries, have either kind of distribution. The number of deliveries might be Poisson (although that is not clear) but the size is something else, which would depend on origin and destination location (e.g., consider power stations vs. a retail outlet). The combined distribution would be very complex.

There are two possible simple remedies. One is to use least squares, which does not require knowledge of distributions provided the Gauss-Markov conditions are met. The key condition, in this context, is that of equality of variance for which empirical diagnostic procedures can be used and combined with weighting. Alternatively, we could depend on the robustness of ML procedures as discussed in Sen and Smith (1995).

**Small-Area Estimation of Origin and Destination Totals**

Since origin and destination total activity levels, $T_{i+} = \sum_j T_{ij}$ and $T_{+j} = \sum_i T_{ij}$, respectively, are at the level of individual origins and destinations, the current literature on small area estimation (e.g., Rao, 2003; Jiang and Lahiri, 2006) can be brought to bear on them. For example, empirical best linear unbiased predictor (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB) estimation and inference methods have been extensively applied to small area estimation.

Generalized linear mixed models, for example, can be used to estimate freight total (weight or value) activity at an origin or a destination as a function of a number of origin- or destination-specific covariates relevant to such an activity. A generic specification might look like: $\log(T_{i+}) = x_i^T \beta + b_i v_i + e_i$ with $x_i$ being the design matrix of area-specific (e.g., number of firms, number of employees, etc.) data, the $b_i$'s known positive constants and the $\beta = (\beta_1, ..., \beta_p)^T$ being the $p \times 1$ vector of regression coefficients. The $v_i$'s are area-specific random effects. A three-year average (likely available at an area agency) can be used as a direct estimator for the for the value of $T_{i+}$ to evaluate the model predictions. Similar actions can be taken at the destination end. In cases where it is not entirely clear as to how these activity levels are defined, we could also use more traditional multiple regression for this purpose – a method which will help identify them, as well as make relevant estimates.

**A Method for Synthesizing Small-Area Origin-Destination Freight Flows**

The methodology we are proposing for obtaining small-area synthetic origin-destination freight flow estimates can be implemented in a number of steps as follows:

- **Step 1:** Obtain origin-destination flows, $N_{ij}^{(L)}$'s, and separation measures, $c_{ij}^{(L)}$'s (e.g., distance, time, etc.), for large (e.g., state to state) areas;
- **Step 2:** Estimate the impedance parameter vector, $\theta$, that corresponds to each separation measure $c_{ij}^{(L)}$, using a ML procedure;
- **Step 3:** Obtain separation measures, $c_{ij}^{(S)}$'s, for small areas (e.g., county to county, TAZ to TAZ);
- **Step 4:** Estimate exogenously for small areas (e.g., counties or TAZs) commodity flow origin and destination activity levels, $T_{i+}^{(S)}$ and $T_{+j}^{(S)}$, respectively using a small-area estimation technique; and
- **Step 5:** Using $\theta$ (from Step 2), $c_{ij}^{(S)}$ (from Step 3), and $T_{i+}^{(S)}$ and $T_{+j}^{(S)}$ (from Step 4), apply an iterative proportional fitting (IPF) procedure, to obtain small-area flow estimates, $T_{ij}^{(S)}$.

Note that in Step 4 above if we exogenously estimate origin and destination activity levels at a small enough geographic scale (e.g., at the TAZ level) then the proposed methodology can be adapted for use in metropolitan freight route planning. In addition, further refinements can be achieved if we disaggregate the freight flows by commodity type and/or industry (assuming data availability) and estimate different models for each commodity.
type and industry group combination. It should also be noted that this procedure allows for the estimation of variance of freight flows at the desired geographic level that can be used to check the reliability of flow estimates and for short-term freight flow forecasting.

Details about the ML procedure used in Step 2 and the IPF procedure used in Step 4 are given in Metaxatos (2004). Both procedures are based on particular algorithmic implementations that facilitate handling of large sparse matrices.

**A Demonstration of the Method Using Publicly Available Data**

In this paper, the above bi-directional small-area estimation (BSAE) methodology is implemented to estimate international trade freight flows from ports of entry to destination counties. Therefore, in Step 4 of the procedure above we do not need to estimate $T_{i+}^{(s)}$ because ports of entry are already refined geographically. We have also simplified the estimation of $T_{+j}^{(s)}$ in Step 4, by simply assigning the total freight tonnage that comes into a state $j$, $N_{+j}^{(L)} = \sum_i N_{ij}^{(L)}$, to counties in the state, based on total employment levels (in all industries) in each county of each state using the U.S. Census Bureau's County Business Patterns data (see Metaxatos, 2004 for details). As a result, an estimate of $T_{+j}^{(s)}$ is obtained.

Regarding freight shipments we used data from the Port Import Export Reporting Service (PIERS) database (see Metaxatos, 2004 for details). A matrix, $N_{ij}^{(L)}$, between 144 origin seaports $i$, and 50 destination states $j$ was developed based on these data. We also used separation measures, $c_{ij}^{(L)}$'s, from each origin seaport $i$ to each destination state $j$ and $c_{ij}^{(s)}$'s from each origin seaport $i$ to each destination county $j$ as discussed in Metaxatos (2004). Avery good model fit was obtained in Step 2 as indicated by formal testing and illustrated in Figure 1.

![Figure 1. Seaport-to-state freight weight flow trip length distribution.](image-url)

All the data items as well as the parameter estimates are now available for Step 5 of the bi-directional small-area estimation (BSAE) procedure to obtain the final estimates of freight flows, $T_{ij}^{(s)}$, between each port $i$ and each county $j$. The reasonableness of the latter estimates was examined as follows:
• Step 1: Computed the freight flow weight flow trip length distribution for seaport-to-state origin-destination pairs using the estimated flows $T_i^j(L)$'s, and separation measures, $c_i^j(L)$'s;
• Step 2: Aggregated the estimated seaport-to-county weight flows, $T_i^j(S)$, into seaport-to-state weight flows, $T_i^j(L)$;
• Step 3: Computed the freight flow trip length distribution for seaport-to-state origin-destination pairs using the newly aggregated flows, and separation measures, $c_i^j(L)$'s; and
• Step 4: Compared the two distributions, either visually or using a more formal statistical method.

As seen in Figure 2 the two distributions are very close. To formally compare the two distributions we used the exact Wilcoxon two-sample test (since the sample size is small, the normal approximation may not be completely accurate, and it is appropriate to compute the exact test). The Wilcoxon statistic (see Agresti, 1992) equals 98.5. The one-sided exact p-value equals 0.323, while the normal approximation yields a one-sided p-value of 0.3276, neither of them significant; the two distributions are practically indistinguishable.

Conclusions

Origin-destination freight flow data that are publicly accessible and of proven quality are not available at a suitable geographic scale for metropolitan freight planning. This paper has presented a method to synthesize such data at desirable geographical levels (e.g., county or TAZ). The proposed BSAE methodology is based on a sound theoretical basis and is also flexible enough to accommodate further refinements, i.e., by industry type, commodity type, etc. The method can be conveniently adopted by MPOs and other practitioners to enhance the traditional battery of transportation planning tools.

We hope the discussion above could motivate additional research in several areas: (a) development of origin-based and destination-based small-area freight production and attraction models that could improve the estimation of small-area marginal totals used in Step 4 of the proposed procedure; (b) development of reliability estimates for the forecasted small-area origin-destination flows, i.e., how accurate the resulted forecasts are; and (c) validation of the proposed small-area methodology using survey data.

References


