

Synthetic Data Generation for Small-Area Demand Forecasting of Freight Flows

Paul Metaxatos

Research Assistant Professor, Urban Transportation Center, and Faculty Associate,
Center for Supply Chain Management and Logistics, University of Illinois at Chicago, USA
Email: pavlos@uic.edu

Abstract

Small area statistics have become increasingly critical for the planning and management of intermodal transportation systems. However, for reasons associated with disclosure of confidential information, data is often released on a fairly coarse geography vis-à-vis a much finer geographical level. This has led to extensive research on small area estimation - i.e., estimation at a more detailed geographical level based on data at a coarser level. Most of this work has been single-area-specific or non-flow data. Freight flows, at a minimum, have origin and destination location specificity, which leads to greater complexity. This paper addresses this issue providing a methodology for small-area estimation of freight flows based on the gravity model. Preliminary empirical findings using publicly available data demonstrate the reasonableness of the method as a freight-planning tool.

Keywords: *small area estimation, gravity model, freight flows*

1. Introduction

The management of intermodal freight transportation systems in the United States is a tremendously complex enterprise requiring 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 (TRB, Special Report 276). This is because freight transportation planning, especially at the metropolitan level, needs flow data at a small geography, say, at 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, Metropolitan Planning Organizations (MPOs) do not have the

resources to undertake expensive freight data collection efforts (TRB, Special Report 288). There is a need, therefore, to develop freight flow data for small areas, that is, at the county level at a minimum.

Small-area estimation methods have been a hot topic in statistics in the last 10 years but they accommodate single-area-specific or non-flow data (e.g. population or economic surveys at a school district, health service area, etc.). Flows, at a minimum, have origins and destinations, which leads to greater complexity. To address this issue, we will provide a methodology for estimating small-area freight flows, which will be then demonstrated on publicly available freight data. The particular application illustrates the distribution of international-trade freight flows within the United States (U.S.) from ports of entry to small-area

destinations (counties). The methodology presented in this paper has a sound theoretical basis and can be adapted for use in metropolitan, statewide and national freight planning.

We could also argue that the discussion below may have implications for freight operations management and supply chain management. This is because industry supply chains are characterized by spatial relationships, which dictate the spatial distribution of commodity flows (Beagan et al, 2007). For example, the spatial organization of distribution networks of a retailer influences the origin-destination patterns of freight flows moving through seaports as part of an international supply chain. These distribution patterns are typically influenced by market areas, for example, locations of distribution facilities close to customer markets. In terms of their importance in freight demand

analysis and forecasting, these critical aspects of the supply chain directly impact the development of commodity flow databases, freight trip generation, and distribution models as well as freight traffic assignment on transportation networks.

The paper can be seen as making contributions to the literature in several areas by: (a) expanding the scope of gravity models to forecasting small-area bi-directional freight flows; (b) proposing a method to extend the statistical literature on location-specific small-area estimation to pair-wise estimation; (c) illustrating the inner workings of maximum likelihood estimation of gravity model parameters that renders the traditional heuristic approaches unnecessary; and (d) providing a method that can be used by planners in the United States and internationally to better manage intermodal transportation systems and supply chains.

Figure 1. The U.S. states of New York and Pennsylvania



2. An Example

Consider two states in the United States (U.S.), e.g., New York and Pennsylvania, as seen in Figure 1. Suppose we know the total flow of commodities in value or weight, T , from New York to Pennsylvania.

Since there are 62 counties in New York and 67 counties in Pennsylvania, a simple-minded approach would be to divide T by 62 and call them estimates of flow between counties. This is too naive. Undoubtedly, the flows are related to some measure of population and economic activity. Call such a measure a_i for the originating state and b_j for the destination state – where i and j index the counties. Then a less naive estimate of flow from i to j would be proportional to $a_i b_j$ and the proportionality can be made into an equality by scaling to add to the total T . However, this too is naive. We know that flows to more difficult to reach destinations would be less. Thus an adjustment is needed and we can estimate the flows to be

$$a_i b_j f_{ij} \quad (1)$$

where f_{ij} reflects the level of difficulty of shipping from i to j and a constant of proportionality is assumed absorbed into one of the factors. But the expression in (1) is the gravity model as explained below. The discussion is akin to that given by Carol and Bevis (see Sen and Smith, 1995).

In expression f_{ij} , (1) was used in a multiplicative way, i.e., f_{ij} is multiplied with the other terms. Intuitively, this is reasonable, since one would expect that doubling activity levels a_i or b_j would double flows irrespective of f_{ij} , so long as it does not change.

3. The Gravity Model

A modern version of the gravity model (see Sen and Smith, 1995) is

$$N_{ij} = T_{ij} + \varepsilon_{ij} \quad (2)$$

where N_{ij} is the flow from i to j , ε_{ij} an error term and $T_{ij} = E(N_{ij})$, the expectation of N_{ij} , is written as

$$T_{ij} = A_i B_j F_{ij} \quad (3)$$

In expression (3), A_i , B_j reflect activity at i and j and F_{ij} reflects the difficulty of getting from i to j or of shipping from i to j .

The gravity model has been intensively studied and has been given a sound theoretical basis. In Sen and Smith (1995), the model has been derived on the basis of a very small number of axioms. Since these axioms are few and intuitively easily believable and also because they are shown to be *both* necessary *and* sufficient, some (including the author) find the theory of the gravity model more complete than the usual theories cited for the logit model. However, because of the relationship between the gravity and logit models, it is most likely that the theory behind the gravity model can be extended to cover the logit models – although this has not occurred so far as we know. Unfortunately all this is not too well known perhaps because the gravity model theory is mathematically rather demanding.

For convenience in estimation, F_{ij} is written, without loss of generality, as

$$F_{ij} = \exp(\phi_{ij}) \quad (4)$$

where $\phi_{ij} = \log(F_{ij})$. It is even more convenient to write

$$\phi_{ij} = \sum_{k=1}^K \theta_k c^k_{ij} \quad (5)$$

where θ_k are parameters to be estimated and c^k_{ij} are variables measuring separation between i and j . From a practical standpoint, this specification also does not lose much generality since it includes polynomials which are known to be dense in the space of continuously differentiable functions. Thus the model we will consider is

$$T_{ij} = A_i B_j \exp\left\{\phi_{ij} = \sum_{k=1}^K \theta_k c^k_{ij}\right\} \quad (6)$$

4. Freight and The Gravity Model

The gravity model has been used before 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). However, most theoretical derivations of the gravity model are confined to

people movements. While it is possible that these theories can be extended to freight (Sen and Smith, 1995; Metaxatos, 2004), for the purposes of this paper, we find the discussion above adequately persuasive.

Note also the close connection between the gravity model and linear programming. If any of the θ s in expression (6) 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 might yield a theoretical basis for the gravity model.

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 (Sen and Srivastava, 1990, p.35). 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).

5. Estimation of Origin/Destination Activity Levels

Since origin and destination 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. 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. We will report on such methods in future papers. In this paper, we will focus instead on the estimation of small-area bi-directional flows when the geography of destination ends is refined.

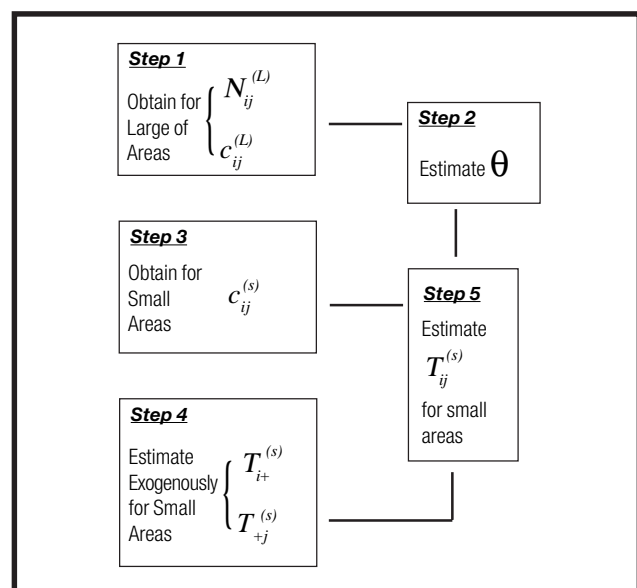
6. An Application

The application presented in this paper focuses on the estimation of (international-trade) freight flows within the United States (U.S.) between U.S. ports of entry and individual U.S. counties. We have borrowed some of the data needed from an earlier study (Metaxatos, 2004) but the methodology below and the results are new.

6.1. 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 is conceptualized in Figure 2. The steps of the procedure are:

Figure 2. Conceptualization of a SAE methodology for origin-destination flows.



- Step 1: Obtain origin-destination flows, $N_{ij}^{(L)}$, 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, θ , that corresponds to each separation measure $c_{ij}^{(L)}$, using the ML procedure presented in Metaxatos (2004);
- Step 3: Obtain separation measures, $c_{ij}^{(s)}$'s, for small areas (e.g., county to county);
- Step 4: Estimate exogenously for small areas commodity flow origin and destination activity levels, $T_{i+}^{(s)}$ and $T_{+j}^{(s)}$, respectively; and
- Step 5: Using θ (from Step 2), $c_{ij}^{(s)}$ (from Step 3), and $T_{i+}^{(s)}$ and $T_{+j}^{(s)}$ (from Step 4), apply the DSF procedure (described in Metaxatos, 2004), a particular algorithmic implementation of an iterative proportional fitting 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 Traffic Analysis Zone (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 given data availability.

6.2 A Few Details About Solution Procedures

For completeness of the presentation, the ML procedure in Step 2 solves the following system of equations:

$$T_{i+} = N_{i+} \quad \forall i \in I \quad (7)$$

$$T_{+j} = N_{+j} \quad \forall j \in J \quad (8)$$

$$\sum_{ij} c_{ij}^k T_{ij} = \sum_{ij} c_{ij}^k N_{ij} \quad \forall k \in K \quad (9)$$

where, I and J are the origin and destination sets, respectively. In addition, the DSF procedure in Step 5 iterates as follows (the index r denotes the iteration number):

$$A_i^{(2r-1)} = O_i / \sum_{j \neq i} B_j^{(2r-2)} F_{ij} \quad \forall i \in I \quad (10)$$

$$B_j^{(2r)} = D_j / \sum_{i=1}^I A_i^{(2r-1)} F_{ij} \quad \forall j \in J \quad (11)$$

where $O_i = T_{i+}$, $D_j = T_{+j}$, and F_{ij} is a function of the separation measures c_{ij}^k . Upon convergence (see Sen and Smith, 1995 for a proof of convergence), the values of T_{ij} 's are given by $T_{ij}^{(2r)} = A_i^{(2r-1)} B_j^{(2r)} F_{ij}$.

To illustrate the ML procedure in equations (7) – (9) we have a prepared a hypothetical example. Table 1 shows a spatial distribution of freight flows among 10 areas (A to J). Each iteration of the ML procedure obtains a better estimate of the θ parameter so that the estimated origin-destination flows are similar to the observed ones according to a chi-square-type criterion. In each iteration, for a given θ the DSF procedure, in equations (9) and (10), re-balances the table to the original origin and destination totals.

Table 1. Observed Origin-Destination Freight Flows (000's tons)

From/ To	A	B	C	D	E	F	G	H	I	J	Origin Total
A	2	25	6	66	25	65	15	3	16	12	235
B	2	14	11	50	18	37	26	4	27	9	198
C	1	8	5	21	13	21	13	8	50	6	146
D	5	9	1	247	61	23	3	0	3	15	367
E	8	15	2	276	87	40	5	1	6	22	462
F	7	29	4	218	79	76	9	3	16	25	466
G	5	59	6	143	53	152	14	3	21	21	477
H	5	62	15	162	61	159	36	7	40	30	577
I	4	35	26	123	44	89	64	11	67	22	485
J	1	8	5	21	13	20	13	8	48	6	143
Destination Totals	40	264	81	1327	454	682	198	48	294	168	3556

Table 2 shows a re-balanced table for a given θ . Notice that the origin and destination totals have been preserved so that equations (7) and (8) are satisfied. Cell-to-cell comparisons between Tables 1 and 2 reveal that the particular θ is not optimal yet because there are some major differences between the origin-destination values in the two tables (more formally, we would obtain a non-significant chi-square-type statistic).

Table 2. Estimated Origin-Destination Freight Flows (000's tons) for a fixed θ

From/ To	A	B	C	D	E	F	G	H	I	J	Origin Total
A	2.9	33.8	5.9	67.3	20.7	56.6	18.5	3.3	18.2	7.9	235.0
B	2.1	15.1	12.9	56.9	12.7	29.7	15.6	6.3	40.1	6.7	198.0
C	1.0	12.0	8.8	32.9	8.6	23.6	12.4	5.9	37.5	3.3	146.0
D	5.5	9.4	1.9	284.2	28.5	18.5	3.2	0.7	4.3	10.9	367.0
E	6.3	17.6	2.6	203.3	139.2	40.5	5.9	1.3	6.9	38.4	462.0
F	6.5	17.9	3.1	207.4	103.1	56.9	9.8	2.1	13.3	46.0	466.0
G	4.7	39.6	5.8	149.6	46.0	173.3	18.4	3.3	21.3	14.9	477.0
H	5.7	57.1	16.0	156.5	48.2	154.6	50.4	9.1	58.1	21.5	577.0
I	4.4	51.8	17.0	142.2	37.3	102.0	53.7	9.7	52.8	14.2	485.0
J	1.0	9.7	7.1	26.7	9.6	26.3	10.1	6.5	41.8	4.3	143.0
Destination Totals	40.0	264.0	81.0	1327.0	454.0	682.0	198.0	48.0	294.0	168.0	3556.0

Table 3 shows origin-destination flows that have been estimated using a near optimal θ . Visual or more formal cell-to-cell comparisons between Tables 1 and 3 can verify the proximity of the values. Additional improvements could be achieved by going through the diagnostic steps that are typical in a model-building process. As an aside, the gravity model in equation (3) can be seen in the context of the log-linear family of models (McCullagh and Nelder, 1989). Thus all the battery of tests available for log-linear models can be brought to bear on gravity model estimation.

Table 3. Estimated Origin-Destination Freight Flows (000's tons) for a near optimal θ

From/ To	A	B	C	D	E	F	G	H	I	J	Origin Total
A	2.6	17.4	5.4	87.7	30.0	45.1	13.1	3.2	19.4	11.1	235.0
B	2.2	14.7	4.5	73.9	25.3	38.0	11.0	2.7	16.4	9.4	198.0
C	1.6	10.8	3.3	54.5	18.6	28.0	8.1	2.0	12.1	6.9	146.0
D	4.1	27.2	8.4	137.0	46.9	70.4	20.4	5.0	30.3	17.3	367.0
E	5.2	34.3	10.5	172.4	59.0	88.6	25.7	6.2	38.2	21.8	462.0
F	5.2	34.6	10.6	173.9	59.5	89.4	25.9	6.3	38.5	22.0	466.0
G	5.4	35.4	10.9	178.0	60.9	91.5	26.6	6.4	39.4	22.5	477.0
H	6.5	42.8	13.1	215.3	73.7	110.7	32.1	7.8	47.7	27.3	577.0
I	5.5	36.0	11.0	181.0	61.9	93.0	27.0	6.5	40.1	22.9	485.0
J	1.6	10.6	3.3	53.4	18.3	27.4	8.0	1.9	11.8	6.8	143.0
Destination Totals	40.0	264.0	81.0	1327.0	454.0	682.0	198.0	48.0	294.0	168.0	3556.0

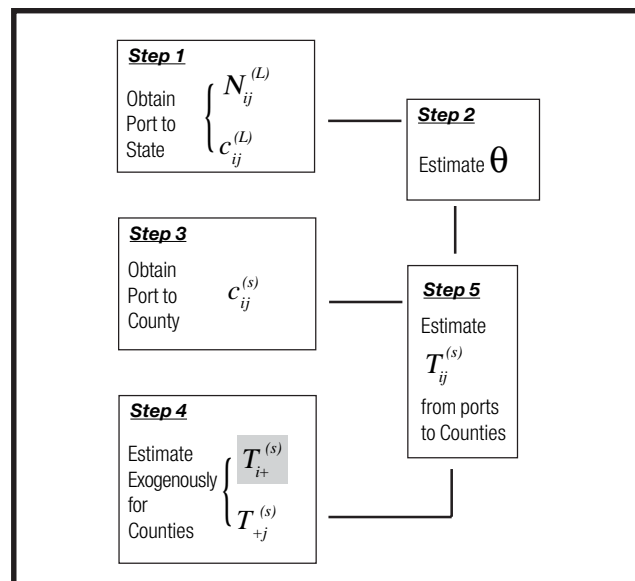
6.3 An Empirical Demonstration of the General Methodology

In this paper, the methodology has been adjusted to the particular application at hand as seen in Figure 3. The objective here is 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. Regarding estimation of $T_{+j}^{(s)}$ in Step 4, we borrowed the method developed in Metaxatos (2004). The method simply assigns 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, <http://www.census.gov/epcd/cbp/view/cbpview.html>). The latter is an annual census of all business establishments with one or more paid employees including information on payroll, employment and industry. As a result, an estimate of $T_{+j}^{(s)}$ is obtained.

6.4 Data Used

The particular application in this paper requires the availability of freight shipments from ports of entry (origins) to destination states. We used a data set made available to us by the Oak Ridge National Laboratory (ORNL). The data are from the Port

Figure 3. Application of the SAE methodology for port-to-county freight flows.



Import Export Reporting Service (PIERS) database. This is a commercial database (see <http://www.piers.com/about/default.asp>) prepared by the U.S. Department of Commerce and offers statistics on global cargo movements transiting seaports in the U. S., Mexico and South America to companies around the globe. A matrix, $N_{ij}^{(L)}$, between 144 origin seaports i , and 50 destination states j with dimensions 144×50 was developed based on this data. Note that a limitation with using PIERS is that reported origins and destinations may be billing addresses rather than shipment points.

In identifying specific types of spatial separation that tend to impede or enhance the likelihood of interactions between ports of entry and destination states, the most obvious types involves physical space as exemplified by travel distance and travel time which are quantifiable in terms of meaningful units of measurement. In this application, we used the two (one for travel distance and one for travel time) separation measures, $c_{ij}^{(L)}$'s from each origin seaport i to each destination state j discussed in Metaxatos (2004). These two separation measures are matrices with dimensions also 144×50 . From the same study we also borrowed two other separation measures (again, one for travel distance and one for travel time), $c_{ij}^{(S)}$'s from each origin seaport i to each destination county j . The last two matrices have dimensions 144×3140 .

6.5 Results

In passenger transportation, a flow unit of 1 (consistent with a Poisson or multinomial distribution assumption) would be reasonable. In the case of freight shipments of goods, a basic unit of flow would appear to be a trainload (for shipments by rail) or a truckload (for shipments by truck). In the absence of mode-specific information as well as information related to the variation in modal size, we experimented with different values and determined an 'optimal' (with regard to providing the best model fit) basic unit of flow of 100,000 pounds (50 short tons). This is not surprising given that the bulk of flows are long-distance shipments that are usually performed by large trucks (currently, the gross weight limit for a 6-axle combination truck is 80,000 pounds (40 short tons) or rail (on a limited scale for the particular data). Interestingly, our results appeared to be quite insensitive to the choice

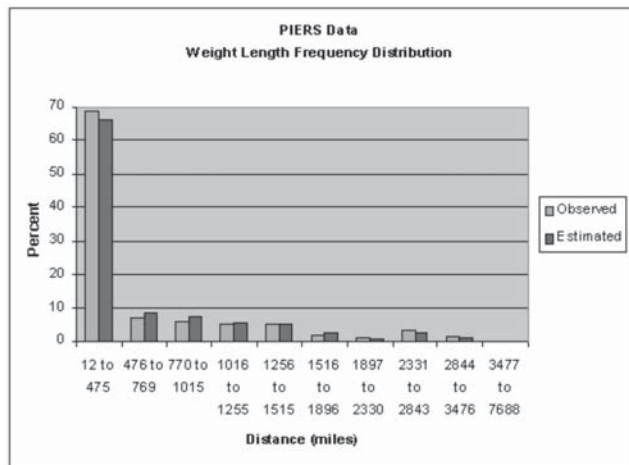
of the flow unit, largely because the flows were inordinately large. This is in agreement with previous work (Sen and Pruthi, 1983).

Note that if we are interested in estimating the impact of goods movement on the transportation network and the environment (e.g, congestion effects, vehicle miles (kilometers) of travel, etc.) we would need to perform a traffic assignment. In that case, we would have to convert the freight flows into truck trips using, for example, the Vehicle Inventory and Use Survey (VIUS) produced by the Census Bureau (U.S. Census Bureau, 2004). In this regard, the previous adjustment is only approximate and can be refined using the weight classes in VIUS. Alternative methods are also available (see, Chin and Hwang, 2006).

The sheer size of the matrices involved in this application called for the use of specialized routines first demonstrated for large-scale freight forecasting in Metaxatos (2004). Both the ML procedure (in Step 2 of the algorithm) and the DSF procedure (in Step 5) run to a tight convergence (see Metaxatos, 2004 for details).

A square-root transformation of travel distance and travel time adequately addressed issues with outliers and resulted in excellent model fit as seen in Figure 4. More formal statistics regarding model fit gave a Chi-square ratio value close to 1, and a Pearson correlation coefficient between observed and estimated cell-to-cell flows with a value of 0.89 (see Metaxatos, 2004 for details).

Figure 4. Seaport-to-state freight weight length distribution.



Estimates of the parameters θ 's corresponding to travel distance and travel time separation measures are, respectively, $\theta_1 = 11.1227$ and $\theta_2 = -16.7447$. The unusual positive sign of the travel distance parameter estimate is due to collinearity between the used impedances, distance and travel time, phenomenon that is well known to adversely affect the sign of parameter estimates. Dropping one of the impedances would bias the parameter estimates left in the model. Given the robustness of the maximum likelihood procedure used in collinearity situations, all available impedance measures were retained (see Sen and Smith, 1995, Chapter 5). After all, the sign of θ_1 would have changed to a negative value had we reparameterized the model and considered instead of travel time, as the first impedance measure, the difference between distance and travel time (in appropriate units).

All the data items as well as the parameters are now available for Step 5 of the small-area estimation (SAE) procedure as follows:

- The freight flow origin activity levels for each seaport i , from the original data since we did not have to refine the geographic scale, i.e., $T_i^{(s)} = \sum_j N_{ij}^{(L)}$;
- The freight flow destination activity levels, $T_{ij}^{(s)}$, for each county j ; and
- The separation measures, $c_{ij}^{(s)}$'s, for travel distance and travel time from each port i to each county j .

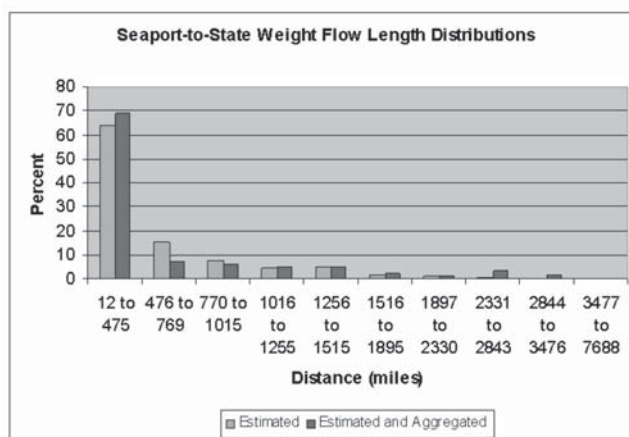
Applying the DSF procedure as described above results in the final estimates of freight flows, $T_{ij}^{(s)}$, between each port i and each county j . A test of reasonableness of the latter estimates is developed as follows:

- **Step 1:** Compute the freight flow length distribution for seaport-to-state origin-destination pairs using the estimated flows $T_{ij}^{(L)}$'s, and separation measures, $c_{ij}^{(L)}$'s;
- **Step 2:** Aggregate the estimated seaport-to-county flows, $T_{ij}^{(s)}$, into seaport-to-state flows;
- **Step 3:** Compute the freight flow length frequency distribution for seaport-to-state origin-destination pairs using the newly aggregated flows, and separation measures, $c_{ij}^{(L)}$'s; and

- **Step 4:** Compare the two distributions, either visually or using a more formal statistical method.

As seen in Figure 5 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.

Figure 5. Comparison of freight weight length distributions.



7. Conclusions

The unavailability of small-scale origin-destination freight flow data that are publicly accessible and of proven quality continues to negatively impact the proper management of goods movements at the national, regional and local levels in the United States. This paper has presented a method to synthesize such data at a small geographical level. It is based on a sound theoretical basis and can be adapted for use in freight planning and management. The framework presented is also flexible enough to accommodate further refinements, i.e., by industry type, commodity type, etc.

The method was demonstrated using available international trade freight flow data and estimated origin-destination freight flows from seaports to counties within the United States. A preliminary verification of the methodology is apparently promising. Understandably, a more robust verification effort would have compared the port-to-county estimated flows with actual data. Regrettably, non-proprietary small-scale data were not conveniently available as of this writing.

However, 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.

Acknowledgements

This research has been made possible with continuous support of the Metropolitan Transportation Research Initiative (METSI) at the University of Illinois at Chicago (UIC) Urban Transportation Center (UTC). We are also grateful to Ashish Sen, Professor Emeritus in the College of Urban Planning and Public Affairs at UIC, who commented on an earlier version of this paper.

References

- Agresti, A. (1992). A Survey of Exact Inference for Contingency Tables. *Statistical Science* 7(1), pp. 131 – 177.
- Ashtakala, B. and Murthy, A.S.N. (1993). Sequential Models To Determine Intercity Commodity Transportation Demand. *Transportation Research. Part A: Policy and Practice* 27(5), pp. 373 – 382.
- Beagan, D., Fischer, M. and Kuppam A. (2007). Quick Response Freight Manual II. Federal Highway Administration. Report No. FHWA-HOP-08-010 EDL No. 14396. Washington, D.C.
- Black, W.R. (1999). Commodity Flow Modeling. *Transportation Research Circular Number E-C011*, pp. 136 – 154.
- Cheu, R.L., Chew, E.P. and Wee, C.L. (2003). Estimating Total Distance for Hauling Import and Export Containers. *Journal of Transportation Engineering* 129(3), pp. 292 – 299.
- Chin, S.M. and Hwang, H. (2006). Converting Freight Flow Information to Truck Volumes. *Preprint # 06-2588*, 2006 Transportation Research Board Annual Meeting, Washington, D.C.
- Erlander, S. and N.F. Stewart (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*, VSP BV, Utrecht, Netherlands.
- Evans, S.P. (1973). A Relationship Between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming. *Transportation Research* 7, pp. 39 – 61.
- Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. *Sociedad de Estadística e Investigación Operativa Test* 15(1), pp. 111 – 999.
- Matsumoto, H. (2007). International Air Network Structures and Air Traffic Density of World Cities. *Transportation Research Part E: Logistics and Transportation Review* 43(3), pp. 269 – 282.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd edition. Chapman and Hall.
- Metaxatos, P. (2004). Estimation and Accuracy of Origin-Destination Highway Freight Weight and Value Flows. *Journal of Transportation and Statistics* 6(2/3), pp. 67 – 80.
- Rao, J.N.K. (2003). *Small Area Estimation*, Wiley, New York.
- Sen, A. and Pruthi, R.K. (1983). Least Squares Calibration of the Gravity Model When Intrazonal Flows Are Unknown. *Environment and Planning A* 15, pp. 1545 – 1550.
- Sen, A. and Srivastava, M. (1990). *Regression Analysis: Theory, Methods and Applications*, Springer-Verlag, New York.
- Sen, A. and Smith, T.E. (1995). *Gravity Models of Spatial Interaction Behavior*, Springer-Verlag, New York.
- Senior, M.L. and Wilson, A.G. (1974). Some Explorations and Syntheses of Linear Programming and Spatial Interaction Models of Residential Location. *Geographical Analysis* 6, pp. 209 – 237.
- Smadi, A. and Maze, T.H. (1996). Statewide Truck Transportation Planning: Methodology and Case Study (With Discussion). *Transportation Research Record* 1522, pp. 55 – 63.
- Transportation Research Board (2003). A Concept for a National Freight Data Program. Committee on Freight Transportation Data. Transportation Research Board Special Report 276. Washington, D.C.
- Transportation Research Board (2007). Metropolitan Travel Forecasting: Current Practice and Future Direction. Committee for Determination of the State of the Practice in Metropolitan Area Travel Forecasting. Transportation Research Board Special Report 288. Washington, D.C.
- U.S. Census Bureau (2004). *Vehicle Inventory and Use Survey*, U.S. Department of Commerce, Washington, D.C.

Paul Metaxatos is a research assistant professor in the University of Illinois at Chicago (UIC) Urban Transportation Center and an affiliated faculty member of the Center for Supply Chain Management and Logistics, also at UIC. He has conducted research in travel demand modeling including freight and passenger demand forecasting, intelligent transportation systems, and transit planning. He holds a doctoral degree in Public Policy Analysis from UIC, a master's degree in Urban Geography, also from UIC, and an undergraduate degree in Rural and Surveying Engineering from the National Technical University of Athens, Greece. Dr. Metaxatos has been a frequent reviewer in academic journals and has also served in numerous professional organizations including the Transportation Research Board, the Institute of Transportation Engineers, the Regional Association International, and the Intelligent Transportation Systems Society of America. Dr. Metaxatos has made numerous publications in academic journals, conference proceedings, technical reports and working papers. More recently, he was a member of the Transportation Systems committee, 10th International Conference on Applications of Advanced Technology in Transportation, Athens, Greece, 2008.