



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

PROCEEDINGS —

Twenty-first Annual Meeting

Theme:

“Transportation Challenges in
A Decade of Change”

October 27-28-29, 1980

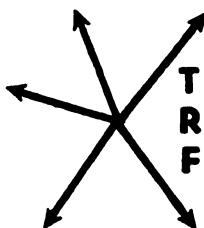
Fairmont Hotel

Philadelphia, Pennsylvania



Volume XXI • Number 1

1980



TRANSPORTATION RESEARCH FORUM

Freight Flow Distribution: A Comparison of Distance Versus Delivered Cost

by Gunter P. Sharp*, H. C. David Yu*, Oscar Adaniya* and Paul S. Jones**

ABSTRACT

A GRAVITY TRADE MODEL is applied to selected commodity specific flows extracted from the Census of Transportation. Instead of the typical distance factor used in the gravity model, there is substituted the delivered cost of the commodity, estimated from production plus transportation costs. This model is used to identify potential markets for new producers by postulating an increase in production at the new origin. The procedure is part of a larger one to identify freight transportation improvements that will stimulate economic development. The procedure is currently being implemented for the Multi-State Transportation Corridor that extends from Brunswick, Ga.-Jacksonville, Fla., to Kansas City, Mo.

I. BACKGROUND

A. Multi-State Transportation Corridor Study

In previous work [Refs. 1, 2] there is described an analytical procedure for jointly identifying economic development opportunities and the transportation facilities needed to assure their economic viability. The approach involves characterizing industry/commodity classes in terms of inputs of raw materials, direct labor, indirect labor, energy, capital, and taxes; and subsequently determining the costs of producing the commodities at a number of prospective locations. The next steps are determining the delivered costs of the products in their respective national markets and predicting marketing viability. Following are steps to identify transportation system improvements. The overall procedure is essentially a screening process, with the economic evaluations and transportation improvements being determined in an

iterative fashion. The procedure is of general application, but is being demonstrated with data from the Multi-State, Multi-Mode Transportation Corridor, an 1100-mile long, 100-mile wide corridor extending from Kansas City, Mo., to Brunswick, Ga.-Jacksonville, Fla. More detailed results are described in [Ref. 3].

B. Role of the Market Share Model

Determining marketing viability is one of the crucial elements in the analytical procedure, for this step leads directly to the question of feasibility. The transportation facilities and service improvements will provide benefits to four different groups:

1. Existing users of transportation facilities within the study area (origins and/or destinations).
2. Existing users of transportation facilities outside the study area who would benefit by using the improved facilities (rerouting flows through the study area).
3. New industries in the study area and expansion of existing industries which can compete effectively in national markets because of the better transportation facilities.
4. Secondary economic sectors in the study area, such as service industries and suppliers to the "export" industries in item 3.

It is anticipated that items 3 and 4 would provide the major justification for any transportation improvements. Accordingly, it is desirable to have a quantitative method for predicting just how much and which types of industries would fit into those categories. The general approach taken is to examine industries at the two- and three-digit SIC level, and to estimate what share of national markets they might capture [Ref. 3]. Economic base analysis is then applied to estimate the secondary effects for item 4 [Ref. 4].

This paper describes the testing of a gravity trade model, which is used to estimate existing flows. The model is then applied to yield information about market areas for new producing areas.

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

**SYSTAN, Inc., P.O. Box U, Los Altos, California.

This research was supported by U.S. DOT Contracts DOT-OS-80612 and DOT-OS-80050.

II. GRAVITY MODELS OF FREIGHT FLOWS

A. Review of Some Previous Work by Others

Black [Ref. 5] was one of the first to demonstrate the use of a gravity model to analyze interregional commodity flows from multiple origins to multiple destinations. Using data from the 1967 U.S. Census of Transportation, he analyzes flows of 80 separate 3-digit TCC classes among the 9 geographic regions of the continental U.S. Basically, his model is of the form:

$$T_{ij}^k = \frac{S_i^k D_j^k F_{ij}^k}{\sum_j D_j^k F_{ij}^k} \quad (1)$$

where

T_{ij}^k = total tons of commodity k produced in region i and shipped to region j .

S_i^k = total shipments in tons of commodity k from region i .

D_j^k = total demand (consumption) in tons of commodity k in region j .

F_{ij}^k = friction factor based on distance d_{ij} .

From a given region i the shipments to region j vary directly with D_j^k , the consumption of commodity k in region j . The friction factor F_{ij}^k varies inversely with distance, so regions that are near a producing zone tend to receive proportionally more shipments. The total shipments and total demand values are assumed known for each region. For the 9 x 9 flow matrices the model produces R^2 values between 0.73 and 0.99. Aggregation of the data to 24 groups does not change these values appreciably. The major problems in drawing conclusions from the study stem from the large (multi-state) zones used and from the validity of R^2 as a test statistic.

O'Sullivan and Ralston [Ref. 6] compare a gravity model with multiple regression and linear programming versions for the 2-digit SIC class with the densest flow matrix, class 20, Food and Animal Products. Flows are based on 1963 and 1967 U.S. Census of Transportation data for 25 metropolitan areas. The linear programming model yields marginally better results than the gravi-

ty model, and both are substantially better than the regression model. The linear programming solution overestimates the major volumes and short distance movements somewhat, at the cost of underestimating smaller flows over longer distances; while the gravity model errors are more scattered as to type. All three models use road distance or a function of it to represent cost, despite the fact that other modes have different circuitries, and hence distances.

Nijkamp [Ref. 7] also compares a gravity model with a linear programming model, using data representing all freight flows among 43 regions in the Netherlands for 1968. The results are inconclusive since he reports only correlation coefficients as goodness-of-fit measures. Pitfield [Ref. 8] makes a similar comparison for highway and rail movements of 30 commodity groups among 134 zones in Great Britain in 1972. Of all the published reports, this one appears to use the most detailed zone structure, and a reasonably detailed commodity classification. The results are inconclusive in selecting among these two types of models. Rather, if one provides a good estimate of the flow matrix for a commodity, then the other is likely to do the same, and conversely. The more important conclusions of the study deal with goodness-of-fit measures, discussed in the following section.

Other relevant work includes that of Wilson and Senior [Ref. 9], who examine the relationships between gravity and linear programming models; and Pitfield [Ref. 10], who calibrates a Poisson-modified transportation problem (allowing sub-optimal routes) to minimize an appropriate goodness-of-fit statistic.

B. Goodness-of-Fit Measures

Much of the difficulty in judging the validity of freight flow models stems from a lack of understanding of the statistical behavior of various goodness-of-fit measures, such as Chi-Square, coefficient of variation, correlation coefficient, etc. [Ref. 8, 11]. For example, the use of the correlation coefficient is erroneous because it measures the degree of linear dependence between two variables, such as actual flow for an O-D pair and the estimated flow. If the actual flow is considered to be a random variable, the estimated flow fails to satisfy certain conditions of independence and normality, since from the gravity model the estimated flow is influenced by the distance between the zones and the total consumption at the demand zone. Therefore, one can expect high, but often meaningless values for the correlation coefficient. This fact has been demonstrated with randomly generated flows [Ref. 11].

Following the work in [Ref. 8, 11], we will use the following two goodness-of-fit measures:

1. Standardized mean absolute error

$$M_{AE}^* = \frac{\sum_i \sum_j |P_{ij} - T_{ij}|}{\sum_i \sum_j T_{ij}} \quad (2)$$

2. Standardized root mean square

$$M_{RS}^* = \frac{[(1/n) \sum_i \sum_j (P_{ij} - T_{ij})^2]^{1/2}}{\sum_i \sum_j T_{ij}/n} \quad (3)$$

where

T_{ij} = observed flow for O-D pair ij

P_{ij} = predicted flow for O-D pair ij

n = total number of O-D pairs.

It should be noted that these are only goodness-of-fit measures, and not statistical measures from which statistically valid inferences can be drawn.

C. Delivered Cost as the Distance Factor

When the friction factor F_{ij}^k is a function of distance in Eqtn. (1), one is making the implicit assumption that distance is the major variable explaining the distribution of freight flows (apart from the magnitudes of supplies and demands). From a microeconomic viewpoint a more plausible variable would be the delivered cost of the commodity [Ref. 3]. Eqtn. (1) then becomes

$$T_{ij}^k = \frac{S_i^k D_j^k \exp(-b^k C_{ij}^k)}{\sum_j D_j^k \exp(-b^k C_{ij}^k)} \quad (4)$$

where

C_{ij}^k = delivered cost of commodity k in zone j from a producer in zone i

b^k = a calibrated parameter for commodity k

Here the delivered cost can be production cost plus transportation cost, or production cost plus transport disutility.

For the application described in the next section, delivered cost was defined as production cost plus a weighted av-

erage of transportation cost by modes, the weights being based on mode split.

D. Evaluating the Production Potential of a New Producing Zone

A typical calibration procedure for the gravity model can be extended to yield information about the marketing viability of a new producing zone. To illustrate this, consider a single iteration through index ij in evaluating Eqtn. (1), using P_{ij}^k for predicted flows:

$$P_{ij}^k = S_i^k \frac{D_j^k F_{ij}^k}{\sum_j D_j^k F_{ij}^k} \quad (5)$$

The predicted flows from each supply zone will always equal the amount of supply:

$$\sum_j P_{ij}^k = S_i^k \text{ for all } i. \quad (6)$$

But the same is not true for demands since $F_{ij}^k \neq 1$. So we expect

$$\sum_i P_{ij}^k \neq D_j^k \text{ for many if not all } j. \quad (7)$$

To satisfy the constraints for demand, an attraction factor A_j^k can be substituted for D_j^k . On the first iteration the A_j^k are set equal to the D_j^k . Thereafter, at each iteration, the values are obtained from:

$$A_j^{k(\text{trial})} = A_j^{k(\text{old})} \frac{D_j^k}{\sum_i P_{ij}^{k(\text{old})}} \quad (8)$$

$$A_j^{k(\text{new})} = a A_j^{k(\text{old})} + (1-a) A_j^{k(\text{trial})} \quad (9)$$

with $0 < a \leq 1$.

At each iteration the new value of A_j^k is used in Eqtn. (5) in place of D_j^k . The weighting factor a is established by experience, with values between 0.3 and 0.8 being typical. At the end of the calibration procedure one then has the parameter b^k and the set of attraction factors A_j^k .

At this point one can add new zones with postulated supplies or increase supplies at existing zones and perform one more iteration. The predicted flows of the new supplies provide information

about which markets can be served and what changes might occur in total demand.

An alternative form of the gravity model can be used to yield slightly different information. If we distribute demand instead of supply, we have

$$P_{ij}^k = D_j^k \frac{S_i^k F_{ij}^k}{\sum_i S_i^k F_{ij}^k} \quad (10)$$

In place of (6) and (7) we have

$$\sum_i P_{ij}^k = D_j^k \quad \text{for all } j \quad (11)$$

and

$$\sum_j P_{ij}^k \neq S_i^k \quad (12)$$

Analogous to A_j^k is a production factor B_i^k for each supply zone. This B_i^k takes the place of S_i^k in Eqtn. (10) and is updated in the same manner as shown in (8) and (9). At the end of the calibration one has b^k and the set of B_i^k . New supplies are then added, with those B_i^k set at their respective S_i^k values and the old supplies retaining their calibrated B_i^k values. After performing one iteration, one has predicted flows from the new supplies. More importantly, one can compare predicted flows with the B_i^k for both old and new suppliers. This form of the model is the one used to obtain the results in the next section.

III. APPLICATION AND RESULTS

A. Input Data

To test the use of delivered cost in a gravity trade model there were used flows for six commodities, shown in Table 1. Flows were obtained for a 120-zone network, described in [Ref. 1]. Flow values were derived from the 1972 U.S. Census of Transportation and National Transportation Plan data, using the procedure described in [Ref. 12].

The estimates of production costs were obtained using data from the Census of Manufacturers. Major inputs for each commodity were determined as follows: direct labor hours, indirect labor hours, capital investment, energy, taxes, and raw material and intermediate inputs by type. Costs were determined by producing region for direct labor, capital investment, energy, taxes, and major raw materials. National average costs were used for the remaining items. Linear production technology was assumed, similar to input-output analysis [Ref. 3]. Transportation costs were estimated with a multi-modal freight network model [Ref. 13]. Costs by different modes were weighted by existing mode splits.

B. Program Outline

The model given by Eqtns. (10)-(12) was programmed in FORTRAN and operated on a Cyber 74 (CDC). Fig. 1 shows an outline of the program logic. The b^k parameter is updated using the formula below [Ref. 14]:

$$b^k(\text{new}) = \frac{b^k(\text{old}) (\text{mean predicted trip cost})}{(\text{mean observed trip cost})} \quad (13)$$

TABLE 1

COMMODITIES USED IN ANALYSIS

Commodity	SIC	Description	Matrix Size	Matrix Density*
1	22	Textile mill products	13 x 26	22 %
2	23	Apparel	13 x 32	14 %
3	24	Lumber and wood products	27 x 49	8 %
4	25	Furniture and fixtures	26 x 30	9 %
5	287	Agricultural chemicals	15 x 42	12 %
6	282	Plastic products	22 x 24	11 %

*Matrix density represents the percentage of non-zero elements in the matrix of O-D flows, with flow values being the sum of flows by mode for each O-D pair.

OUTLINE OF PROGRAM LOGIC

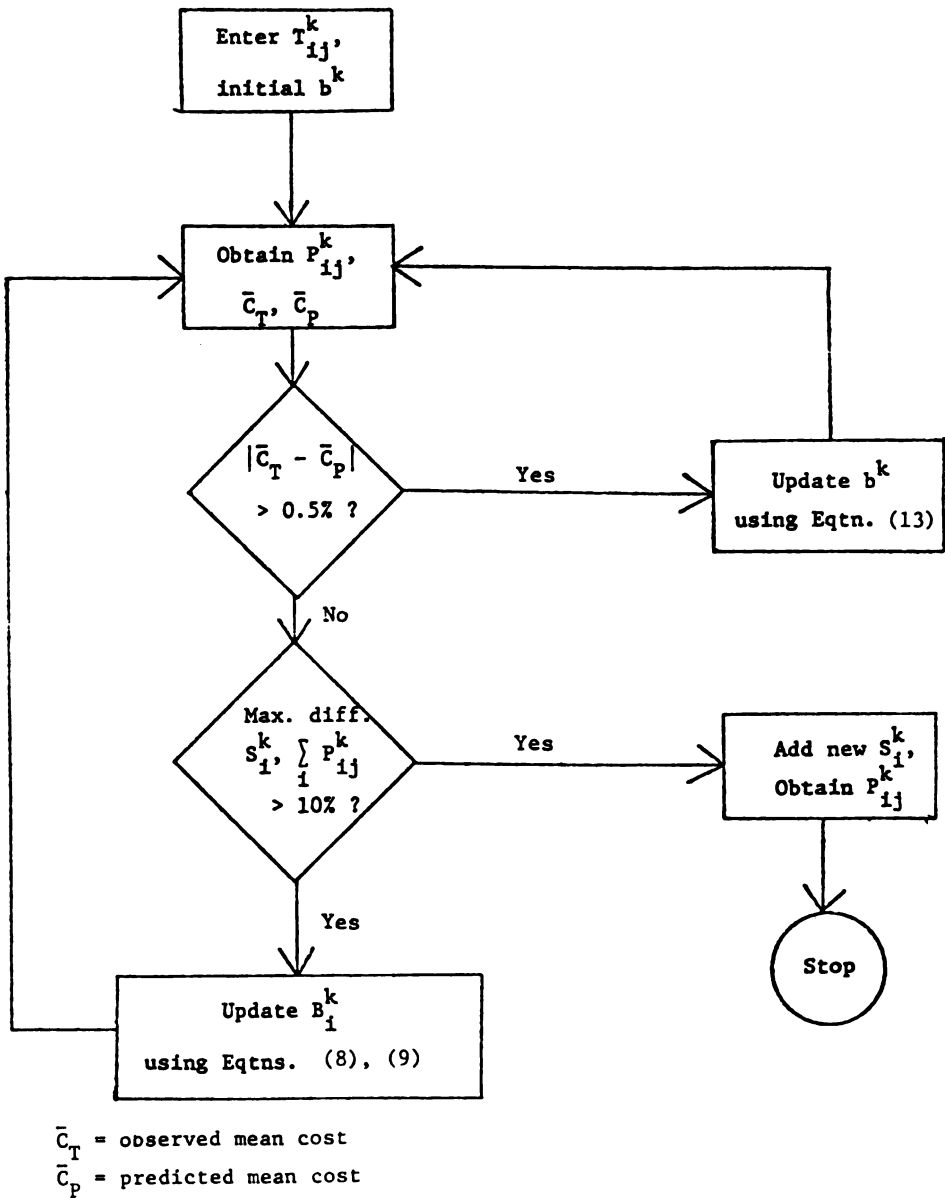


FIGURE 1

The calibration terminated after satisfying two criteria:

1. Difference between mean predicted and observed trip cost was less than 0.5%.

2. Maximum difference between S_i^k and $\sum_j P_{ij}^k$ was less than 10%.

Two functional forms of the friction factor were used:

Form 1:

$$F_{ij}^k = \exp(-b^k d_{ij})$$

distance version (14a)

or $F_{ij}^k = \exp(-b^k C_{ij}^k)$

delivered cost version (14b)

and Form 2:

$$F_{ij}^k = (d_{ij})^{-b^k}$$

distance version (15a)

or $F_{ij}^k = (C_{ij}^k)^{-b^k}$

delivered cost version (15b)

C. Results for Existing Flows

The test results for existing flows

were quite surprising. The distance versions of the friction factor resulted in better overall goodness-of-fit measures than the delivered cost versions. This was true for the first five of the six commodities, as shown in Table 2. The statistic R^2 is included here for comparison with other studies, but should not be relied on for meaningful interpretation. No general conclusions can be drawn about the functional form of the friction factor.

More detailed examination of the analysis data revealed another quite unexpected result: The model with worse aggregate goodness-of-fit measures had production factors B_i^k that were closer to their supplies S_i^k . This phenomenon is shown in Table 3. The delivered cost version consistently yielded production factors that were closer to the S_i^k values

TABLE 2

AGGREGATE GOODNESS-OF-FIT MEASURES FOR EXISTING FLOWS

Commodity SIC class	1 22	2 23	3 24	4 25	5 287	6 282
Form 1						
distance R_{MS}^*	1.17	1.72	2.20	2.26	1.84	3.00
version M_{AE}^*	.66	.75	.67	.84	.73	.96
R^2	.78*	.77	.92	.83	.93	.65
delivered R_{MS}^*	1.53	2.67	2.47	4.29	4.34	2.00
cost M_{AE}^*	.83	1.00	.73	1.43	1.10	.85
version R^2	.62	.45	.90	.37	.59	.85
Form 2						
distance R_{MS}^*	1.28	1.73	1.99	2.25	1.85	3.09
version M_{AE}^*	.71	.75	.57	.81	.71	.93
R^2	.73	.77	.93	.83	.93	.63
delivered R_{MS}^*	1.67	2.58	2.42	4.29	4.46	1.95
cost M_{AE}^*	.89	1.01	.72	1.43	1.11	.83
version R^2	.55	.49	.90	.37	.57	.85

* R^2 is obtained from Total Sum of Squares less Error Sum of Squares, since this is a nonlinear estimation model.

TABLE 3

**PRODUCTION FACTORS VERSUS SUPPLIES,
STANDARDIZED MEAN ABSOLUTE ERROR***

Commodity	1	2	3	4	5	6
Form 1, distance version	.092	.002	.022	.004	.024	.005
deliv. cost version	.023	.000	.000	.000	.008	.044
Form 2, distance version	.012	.010	.016	.021	.015	.009
deliv. cost version	.000	.000	.000	.000	.000	.112

* Goodness-of-fit measure computed from

$$\frac{\sum_1^k |S_1^k - B_1^k|}{(\text{no. of supply zones}) \left(\sum_1^k S_1^k \right)}$$

for the first five commodities. In fact, there were eight (out of twelve cases) when the B_1^k were exactly equal to the S_1^k for the delivered cost model version. For the sixth commodity, the only one for which delivered costs yielded better aggregate fits, the distance versions gave B_1^k values closer to the S_1^k values.

It is somewhat difficult to interpret the above findings. Checking and recheck of the runs confirmed the results. On the surface, it would indicate that the distance versions of the gravity model generally give better overall data fits but require more manipulations of the calibration parameters, which consist of the exponent b^k in the friction factor and the production factors B_1^k . The delivered cost versions, on the other hand, are often able to give quite good fits with no adjustment of the B_1^k . In other words, in many cases the delivered cost is a better explanatory variable for the flow distribution than distance by itself.

The above tends to favor the delivered cost version for evaluating new producing zones. In order to use the distance version for predicting flows from new zones, one must know the economic characteristics of that and all other producing zones affect the production factor B_1^k for the new zone. But there does not seem to be any convenient way to determine how the B_1^k should vary from the supply S_1^k . The delivered cost version requires little or no adjustment of the S_1^k , and hence it appears to be a more appropriate model.

It is perhaps premature to draw such conclusions without examining the re-

sponse surface of the aggregate goodness-of-fit measures with respect to b^k and the B_1^k . The computer program tends to spend more time adjusting b^k . In one run the value of b^k changed from 1.0 to 0.1 when the convergence criterion for mean predicted and observed trip cost was reduced from 1% to 0.5%. Also, the considerable input data on production costs need to be scrutinized.

Generally speaking, both the distance and delivered cost models perform well when compared with other studies. The M_{AE}^* and R_{MS}^* values are generally less than those reported by Pitfield [Ref. 8]. Again, it is inappropriate to draw conclusions here, since the latter study used a more detailed zone structure. The R^2 values, although meaningless, range from 0.63 to 0.92 for the distance versions and from 0.37 to 0.93 for the delivered cost versions. CPU times ranged from 5 to 40 seconds per run, with 10 seconds being typical (Cyber 74 CDC).
D. Results for Production Potential

The production potential for four new zones are evaluated for the first commodity using the procedure described in Section II.D. The S_1^k (and B_1^k) values for each of the new zones was set equal to the average supply for existing producing zones. Since no other information was available concerning the hypothesized production amounts, the values were set in this manner. The predicted shipments after one more iteration are then compared with the calibrated production parameters B_1^k , as shown in Table 4.

The ratios tend to be less than 1.0 because the total supply, with the new zones, exceeds the existing demand.

TABLE 4

**RATIOS OF PREDICTED SHIPMENT FROM ZONES
VERSUS PRODUCTION FACTORS, COMMODITY 1 (SIC 22)**

Supply Zone	Form 1		Form 2	
	distance v.	del. cost v.	distance v.	del. cost v.
1	.43	.38	.63	.77
2	.66	.65	.83	.77
3	.61	.55	.71	.77
4	1.28	1.30	.53	.76
5	1.65	2.01	.46	.76
6	1.23	.72	.38	.75
7	.87	.54	.82	.77
8	.88	.52	.78	.77
9	.90	1.03	1.10	.77
10	.78	.58	.64	.76
11	.85	.91	.84	.76
12	.75	.55	.85	.77
13	.66	.43	.70	.76
14*	.52	.48	.48	.77
15*	.48	.46	.45	.77
16*	.48	.46	.46	.77
17*	.41	.41	.41	.77

* New Producing zone

With form 1 the ratios for the distance version are similar to those for the delivered cost version, as applied to existing zones; there is close agreement between the two versions when applied to the four new zones, numbers 14-17. The ratios for the new zones are lower than for the existing producing regions, implying that it would be more difficult, all other factors being equal, to ship the designated quantities from those than the corresponding quantities from the existing zones. Such a conclusion tempers the more optimistic ones in [Ref. 2, 3]. Since existing flows are the result of many factors, however, it would be wise to experiment with different values of S_k for the new zones.

The distance version of form 2 yields ratios that vary considerably from form 1 for existing zones. For the new zones the results are similar to those of form 1. The anomalous result for the delivered cost version (all ratios in the range 0.75-0.77) has not been explained yet.

Results for the other five commodities were similar to those for the first. The

above analysis indicates the relative strengths of the new producing regions and identifies the specific market areas that can be served. The actual shares of these markets will be determined from more detailed market studies. The advantage of the freight flow model is that it serves as an effective screening device for the numerous commodity-producing zone-market zone combinations that are possible.

IV. CONCLUSIONS

It is apparent that gravity models with either distance or delivered cost in the friction factor can yield adequate predictions of existing interregional commodity flows. The distance versions provide better aggregate goodness-of-fit measures for five of six commodities tested, while in the delivered cost versions less manipulation of the production factors is needed. This fact indicates that the delivered cost version is more easily transferred and applied to new producing zones. A typical application

would require one calibration of the exponent in the friction factor, using data for existing flows. Extension of the model to evaluate the production potential of new producing zones was also demonstrated. Future research is indicated in the areas of the response surface of the goodness-of-fit measures, and in the estimation of actual market shares.

REFERENCES

1. Jones, P. S. and G. P. Sharp, "Multi-mode Intercity Freight Transportation Planning for Underdeveloped Regions," *Proc. 18th Mtg. Trans. Res. Forum*, 1977, pp. 523-531.
2. Jones, P. S. and G. P. Sharp, "An Analytical Process for Coupling Economic Development with Multi-Mode and Intermodal Transportation Improvements," *Trans. Res. Rec.* No. 686, 1978.
3. Jones, P. S., Ed., *Procedures for Multi-State Multi-Mode Analysis: First Year's Research, Final Report*, Report No. DOT-OS-60512-8, Georgia Inst. of Tech., School of ISyE, Atlanta, Georgia, Dec. 1977.
4. Sharp, G. P., Ed., *Procedures for Multi-State Multi-Mode Analysis: Second Year's Research*, Report No. DOT-OS-80050-9, Georgia Inst. of Tech., School of ISyE, Atlanta, Georgia, May 1979.
5. Black, W. R., "Interregional Commodity Flows: Some Experiments with the Gravity Model," *Jrl. Regional Sci.* No. 12-1, 1972, pp. 107-118.
6. O'Sullivan, P., and B. Ralston, "Forecasting Intercity Commodity Transport in the U.S.A.," *Regional Studies*, No. 8, 1974, pp. 191-195.
7. Nijkamp, P., "Reflections on Gravity and Entropy Models," *Reg. Sci. and Urb. Econ.*, No. 5, 1975, pp. 203-225.
8. Pitfield, D. E., "Freight Distribution Model Predictions Compared: A Test of Hypotheses," *Environment and Planning A*, No. 10, 1978, pp. 813-836.
9. Wilson, A. G., and M. L. Senior, "Some Relationships Between Entropy Maximizing Models, Mathematical Programming Models, and Their Duals," *Jrl. Regional Sci.* No. 14-2, 1974, pp. 207-215.
10. Pitfield, D. E., "Sub-Optimality in Freight Distribution," *Trans. Res.*, No. 12, pp. 403-409.
11. Wilson, S. R., "Statistical Notes on the Evaluation of Calibrated Gravity Models," *Trans. Res.*, No. 10, pp. 343-345.
12. Mullens, M. A., G. P. Sharp, and P. S. Jones, "Development of a 47-Commodity Flow Table for BEA Zones," *Proc. 19th Mtg. Trans. Res. Forum*, 1978.
13. Sharp, G. P., "A Multi-Commodity, Intermodal Transportation Network Model," *Proc. 20th Mtg. Trans. Res. Forum*, 1979, pp. 399-407.
14. Williams, I., "A Comparison of Some Calibration Techniques for Doubly Constrained Models with an Exponential Cost Function," *Trans. Res.*, No. 10, pp. 91-104.