Forecasting freight demand

by
PAUL O. ROBERTS
Massachusetts Institute of Technology, Massachusetts

FREIGHT FORECASTING: A NEGLECTED AREA

Freight forecasting is an area that appears to have received considerably less attention than it deserves. The literature available describes a range of techniques that have been applied over the years, none of which seems to have worked extremely well. There are no routinely accepted and applied approaches. Even the "gravity model", which is the closest thing to a standard technique, is used either out of total naivety or complete frustration over the lack of suitable alternatives. One is forced to conclude that quantitative forecasting techniques for freight are not well developed, at least operationally, in comparison to those used for urban transportation planning.

Two possible reasons for this lack of development can be advanced. The first is the complexity of dealing with freight rather than passengers. The second is the lack of a consistent and comprehensive data base for use in development. Let us briefly examine the arguments for each.

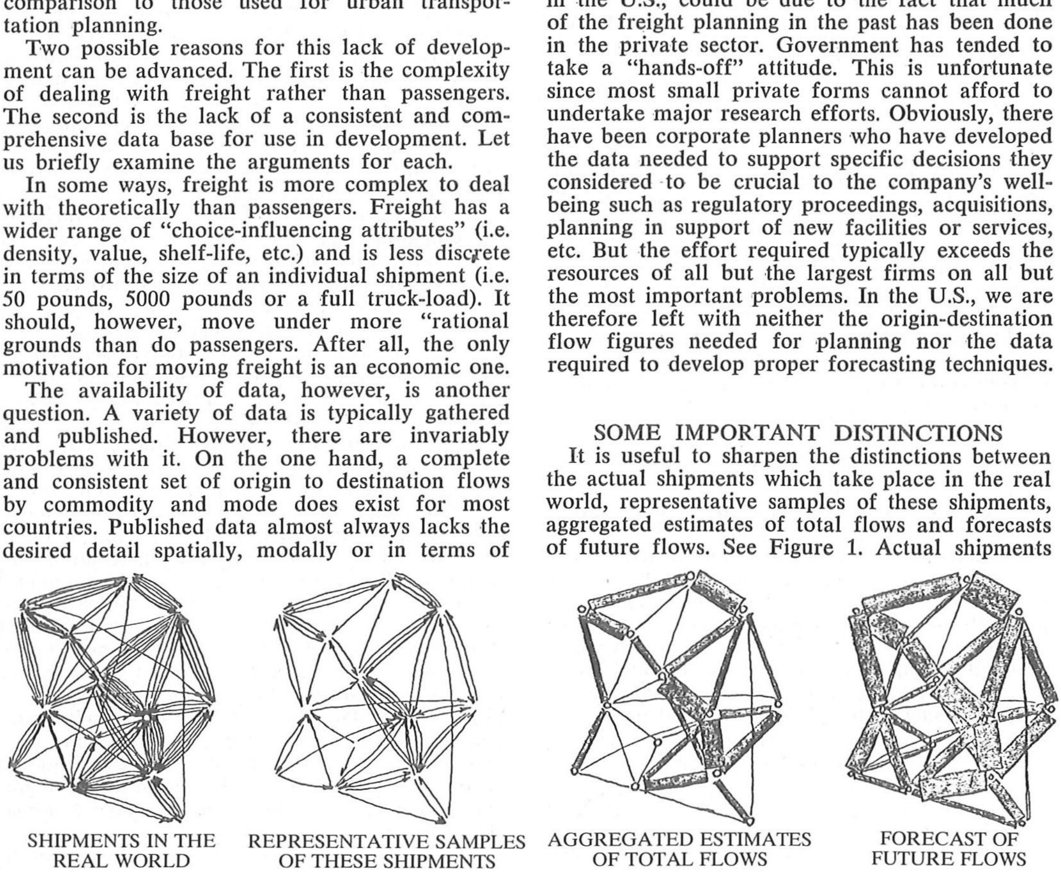
In some ways, freight is more complex to deal with theoretically than passengers. Freight has a wider range of "choice-influencing attributes" (i.e. density, value, shelf-life, etc.) and is less discrete in terms of the size of an individual shipment (i.e. 50 pounds, 5000 pounds or a full truck-load). It should, however, move under more "rational grounds than do passengers. After all, the only motivation for moving freight is an economic one.

The availability of data, however, is another question. A variety of data is typically gathered and published. However, there are invariably problems with it. On the one hand, a complete and consistent set of origin to destination flows by commodity and mode does exist for most countries. Published data almost always lacks the desired detail spatially, modally or in terms of commodity disaggregation. On the other hand, a representative sample of individual shipments from which estimates of the origin to destination flows could be prepared either does not exist or the holder of such data is reluctant to release it because its disclosure might reveal the operations of individual firms.

Why this data gap has been allowed to exist is hard to explain. Obviously, everyone's data needs are not the same. Developing a complete and consistent set of origin to destination flows at a level of detail that would be generally satisfactory might well be considered to be too expensive for most government agency planning budgets. Another explanation for this lack of data in the U.S., could be due to the fact that much of the freight planning in the past has been done in the private sector. Government has tended to take a "hands-off" attitude. This is unfortunate since most small private forms cannot afford to undertake major research efforts. Obviously, there have been corporate planners who have developed the data needed to support specific decisions they considered to be crucial to the company's well-being such as regulatory proceedings, acquisitions, planning in support of new facilities or services, etc. But the effort required typically exceeds the resources of all but the largest firms on all but the most important problems. In the U.S., we are therefore left with neither the origin-destination flow figures needed for planning nor the data required to develop proper forecasting techniques.

SOME IMPORTANT DISTINCTIONS

It is useful to sharpen the distinctions between the actual shipments which take place in the real world, representative samples of these shipments, aggregated estimates of total flows and forecasts of future flows. See Figure 1. Actual shipments



SHIPMENTS IN THE REAL WORLD — REPRESENTATIVE SAMPLES OF THESE SHIPMENTS — AGGREGATED ESTIMATES OF TOTAL FLOWS — FORECAST OF FUTURE FLOWS

Figure 1 - Four types of shipment information

can never be known without some kind of sample. There may, in any particular case, be more than one sample. Samples are typically summed and expanded by an appropriate factor to produce an aggregate estimate of total flows. Aggregate estimates can be further aggregated to provide summary demand statistics. For example, total ton-miles for the U.S. as a whole is one such super-aggregated statistic. Aggregate estimates then, have their origins in disaggregate samples. It is relatively easy to sum disaggregate samples to obtain aggregate estimates of the whole. It is much more difficult to "disaggregate" an aggregate estimate since information has actually been given up in the process of aggregation.

Forecasts of future demand are different in several important ways from estimates of existing origin to destination flows. There is only one past. It is a matter of certain historical record, if the record was saved. By contrast, there are literally an infinitude of possible futures. The one that will occur is more or less uncertain. It depends, in part, on the chance occurences that make up the environment within which the transport system is located and in part on the choices which an individual decision maker exercises. Forecasting future flows always involves the use of a model which incorporates these chance and choice elements. The model may be either quantitative or qualitative. It could involve simple extrapolation or more complex causal relationships. It may be assumed to be either determinstic or probablistic. It may involve digital or analog computation or no computation at all. However, there is always a model either explicitly or implicitly involved since the future, if it is interesting at all, always involves changes from the present.

Since we are concerned here with forecasting freight demand we must, as a consequence, be concerned with the nature of various models for accomplishing this purpose. We must also be concerned with the status of the data both disaggregate and aggregate since it serves both as a starting point for estimates of future flows and as the only source of quantitative information on causal responses to changes in the process. Thus, models and data are closely tied. Ultimately, one cannot know whether a model performs properly without comparing its predictions against observed changes in the data.

FACTORS AFFECTING MODELS OF FREIGHT DEMAND

Freight demand forecasting has a wide variety of uses, ranging from providing estimates of future inputs to the transport sector to determining the volume of flow on a particular link as a function of price. The exact uses will depend importantly on who the user is and what he is using the forecast for. For some uses, the precision required is not great. For others, as much accuracy is desired as can be obtained. Obviously, the type of model to be used will depend upon such factors as degree of aggregation, policy responsiveness and requirements for accuracy.

1) **Aggregation** — There are at least four important dimensions over which aggregation can be performed. These are: time, space, commodities and modes. See Figure 2. Most models deal with time only in the most basic way. That is, demand is expressed as a rate per unit of time. Time is not explicitly considered as a separate variable as it is, for example, in a Monte Carlo simulation of a time dependent process such as queuing. It could be treated but for most demand forecasts, the average rate of flow per day, per week, or per year is satisfactory. Where seasonality of flow is important, a separate estimate can be made by seasons. However, this is a complication that is difficult to handle in existing models.

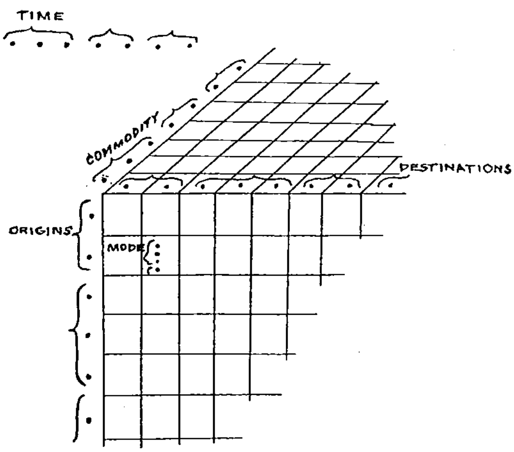


Figure 2 - Common dimensions of aggregation

Spatial aggregation is also common in demand forecasting though the need to preserve spatial detail is widely recognized. Where the origin and destination areas are large, there is also a question on the inclusion of intra-regional flows in addition to the interregional flow. The ability to preserve spatial detail in the forecasts depends importantly on the spatial detail which exists in the base data used in developing the model and the availability of values for independent variables in compatible detail. In the U.S., some data for independent variables are available from published statistics at various levels ranging progressively from the county, SMSA, state and region to the nation as a whole. Unfortunately, agregate flow statistics are available between only a selected set of major metropolitan areas and then for only a portion of the commodities in the total economy in relatively aggregate form.

The most pervasive aggregation occurs over commodities. The great diversity in attributes between commodities, even those within rather detailed commodity categories, and the importance of these attributes for the selection of mode tend to make even moderate aggregation unfortunate. There are a great many commodity coding systems in use throughout the world. In the U.S., the Standard Industrial Classification (SIC) and the compatible Standard Transportation Commodity Classification (STCC) are widely used. Establishments tend to be classified at 4 digits, particular commodities at 5 digits and commodities in various forms of packaging at 6 and 7 digits. Since freight rates tend to be quoted at the 6 and 7 digit level, significant disaggregation must exist if the ability to make a modal choice is to be preserved.

Aggregation over modes is less common probably because of the rather large distinctions between the costs and the services rendered. Rarely, for example, does one see a single figure for ton-miles over the whole economy while ton-miles by rail, by truck, or by air are common. There is, however, considerable aggregation over sub modes. Piggyback is frequently undistinguished from rail carload, which is also typically not differentiated from rail unit train. Likewise, the various categories of private and forhire trucking are frequently combined. A major distinction between various modal service offerings is the minimum shipment size required to qualify for a specific set of tariff charges. For example, LTL truck shipments are almost a different mode from truckload shipments.

A great many of the differences found in the various models and modelling approaches can be accounted for by different levels and types of aggregation. Depending on the level of detail required for addressing the questions of interest, the models at higher level of aggregation may not be useful at all.

2) **Policy Responsiveness** — A second major factor of interest in developing models of freight demand is policy responsiveness. Policy responsiveness becomes important whenever the choices open to a decision maker become important. By contrast, the chance elements are not of particular interest unless particular choices are impacted differentially. Obviously, the policy responsiveness of a particular model is affected by the variables incorporated into the model. It would, for example be difficult for the demand for high speed service to be forecast by a model unless the model incorporates travel time for each of the modes. Each policy to be evaluated by the model must be expressed in terms of the variables in the model. This would become almost impossibly complicated if it were not that almost all transport policy can be expressed in terms of a relatively small set of level of service variables. For freight, this set might typically consist of:

waiting time (or schedule frequency)
trip time
time reliability
probability of loss and damage
minimum shipment size
transport charges
other costs

Factors such as availability of special handling, expediting, environmental controls, transit privileges, enroute tracing, facilitated claims processing, etc. could also be important, but they are not as general as the list above.

The policy responsiveness of the models can be enhanced by proper model design. That is, some model features work in favor of policy responsiveness, others against it. Clearly, what is desired is a model which for a variety of decision makers and/or types of decisions can reflect the choices that would be exercised by the decision maker as he faces different level of service combinations reflecting the policies of interest.

3) **Accuracy** — Model acuracy is the third major factor. Clearly, if policy responsiveness is important, the accuracy required is that necessary to discriminate between policies. This will vary depending upon the nature of the policy. It could require more or less aggregation over time, space, commodities or modes. If sampling techniques are used, explicit criteria could be developed to produce confidence limts for any forecast. The modelling methodology will also be important. This will be discussed in greater detail below.

CAN GENERAL PURPOSE MODELS BE DEVELOPED

So far, we have not described a particular model or class of models, but have rather described factors affecting all models regardless or who was using them or the questions that were being addressed. There is some question as to how long maintaining this generality is useful. Is it possible, for example, to design a general purpose freight demand model that can be used by all parties on all types of problems? The answer is probably no, but the degree to which generality can be built into any model will obviously govern its usefulness.

The purpose of a freight travel demand model is clearly to allow the forecasting of the volume of travel of a particular commodity that will move between a given origin and destination by a given modal service offering over a given interval of time. A demand model does not take a particular point of view. There is no inherent difference between the demand model used by a carrier to determine the volume of travel that would take place if a specific change in service offering were made and the demand model used by the federal government to understand the same problem. Thus, one demand model should work for all types of users.

The same question arises as to issues. Can all types of policy questions be asked and answered with the same model? The answer is probably mode dependent on the factors of aggregation, policy responsiveness and accuracy mentioned above than it is on the issues to be addressed. There is no inherent reason why issues concerning labor, technology, pricing, regulation, capital investment, equipment utilization, or the like cannot be addressed using the same demand model as long as the issue depends upon the estimates of demand for a particular commodity or set of commodities between given origins and destinations by individual modal service offerings over a given interval of time. Some issues could require an ability to discriminate between impacts by type of shipper, but this depends on the features of the modelling approach. It is possible to select modelling approaches that can handle this routinely. If so, it would appear that a more or less general purpose model could be developed which was broadly policy responsive subject to the factors of aggregation and accuracy.

DESIRABLE MODEL FEATURES

A careful review of the literature, such as that performed by Terziev [1], reveals a series of models developed over the years at different levels of aggregation and accuracy. The models reviewed were not very policy responsive and for the most part performed poorly, even for their stated purpose. The question, of course, is whether this is an inherent property of all freight travel demand models or whether the problems lie with the features of the models constructed to date.

Our hypothesis is that the problems lie with the features of previous modelling approaches and that the adoption of a set of carefully thought out modelling features could greatly improve the generality as well as the utility of present approaches. These features then can be stated as criteria for constructing the desired travel demand models. These criteria are:

1. Work at a disaggregate level.
2. Model the behaviour of individual decision making units.
3. Use the shipment as the basic element to be modelled.
4. Base the model at the destination.
5. Determine model parameters empirically.
6. Formulate the model to use generalized attributes.
7. Base the computation scheme on forecastable data sources.

Each will be discussed in more detail in the sections which follow.

1) **Work at a Disaggregate Level** — Since most forecasts must be reported and used at an aggregate level, a model that works at an equivalent level of aggregation would appear to be the simplest to use. This is undoubtedly true, but a model formulated at a higher level of aggregation is not usable at a more disaggregate level. For example, a model for the United States of ton-miles of transportation by truck as a function of GNP is simpler to use than a model of state-to-state truck flows based on state-level economic indicators. However, the state-to-state model can be used to generate ton-miles between specific states, ton-miles by region, and ton-miles by state of origin or destination. Importantly, the results can also be summed to produce total ton-miles by truck for the United States as a whole. The more disaggregate model is, therefore, more flexible and can be used for purposes which the higher level model cannot be used. This generality is gained at the expense of being slightly more difficult to use.

In general, models formulated at lower levels of aggregation have more general utility, since they can be combined in more ways. The results can always be aggregated to obtain the same results as those of the more aggregate models. By summing the disaggregate units, the information content inherent in the disaggregate units is not lost.

Another feature of using the mode disaggregate models is their increased policy sensitivity. Policies which apply differentially to the individual sub-units can be analyzed. Changes in the environment which impact the sub-units differentially can also be handled. Thus, a policy such as the imposition of differential user charges in the various states could be handled by the disaggregated state-to-state model described above, whereas the aggregated U.S. model could not incorporate such a change.

The disadvantages of a more disaggregate model are that there are more inputs to be dealt with, there is more computation to obtain disaggregate forecasts, and there must be a scheme for aggregation, which in some instances might be quite complicated. In general, however, the advantages of using more disaggregate approaches seem to outweigh the disadvantages. The models are more flexible. That is, they can handle a wider range of policies and a wider range of environmental change. It is not necessary to know at the time the model is being developed the exact policies or exact environments as carefully as it is for a more aggregated model. The ability to reaggregate according to a new aggregation scheme is an extremely valuable asset to a set of models since it automatically increases their generality.

2) **Model the Behavior of Individual Decision Making Units** — If an individual decision making unit can be identified, it is much easier to understand this individual's point of view. In many cases, it will be possible to understand his objectives and to identify the choice variables which he can exercise. For freight modelling, this individual will be either a shipper or a receiver of freight. It should be possible, therefore, to understand the costs which he faces, and to hypothesize a cost function in general terms which approximates his view of the world. This should greatly improve freight demand models, because it removes ambiguity from the shipping process. Instead of a generalized cost function representing impedence crudely defined, we can have a cost function which incorporates those elements typically faced by the decision maker.

Likewise, the choices that are available to the decision maker can be more easily identified. By knowing the choices that are available and if the costs can be developed for each, a model of the decision making process can be formulated. There is now a theoretical basis using consumer theory for formulating a utility, or cost function, and the theory of consumer choice can be applied in the formulation of appropriate models. It should no longer be necessary to fall back on the gravity model because there is nothing better.

There are some disadvantages to attempting to model the individual decision making units. It may no longer be possible to simply sum the component parts, as it was in the case of state-to-state flows, to obtain the national flow. We must now have an explicit scheme for aggregation. We may, for example, want to use a representative sample of the population as a whole in which the choices made by each individual are recorded, and the final results factored up to represent the total universe.

Working with individual decision making units carries with it one major advantage. The advantage is "transferability". This means that once a model is available for one part of the country, it can be used in other parts as well. A model that works for any decision making unit can be transferred from place to place as long as the individual decision making units do not change in character; that is, where the changes can be described by inputs to the model. For this to be the case, a model must be based on attributes of the individual, and not the individuals themselves. There are, obviously, considerable economies in calibrating such a model once and for all.

Another extremely important feature of the use of individual decisionmaking units as the basis for modelling is the ability to determine the impacts of a specific policy on certain definable groups of individuals. For example, if various individuals can be identified as belonging to a particular group, i.e. a given industry, firm size, region, etc., then the impacts on this particular group can be determined by merely isolating the individual observations and observing their behavior before and after the policy application. This feature is extremely important in issue-oriented questions. There is frequently a need for determining the groups that will be impacted as a result of a specific policy-changes in user taxes, pollution, etc. Models based on the behavior of individual decision making units are amenable to tracing the impacts characteristic of issue-oriented policies. Aggregation tends to lose this ability.

3) **Use the Shipment as the Basic Element to be Modelled** — Standard practice in passenger demand modelling is to use the person-trip as the element to be modelled. In the case of freight, there are a number of possible choices for the basic element to be modelled. These include vehicle trips by mode, ton miles by mode, tons irrespective of mode and shipments. Commodity type may or may not be explicitly considered. The arguments for one as opposed to the other must be developed in more detail for the advantages of using shipments to be apparent.

Modelling truck trips, barge trips, rail trips or air trips directly is filled with problems. Train length and barge tows can vary in size. Equipment with different playloads can be used. Imbalances in flows can lead to different load factors in different directions which cannot be represented easily. Intermodal trips cannot be handled without special treatment. Commodity distinctions are also difficult, if not impossible. In spite of these apparent disadvantages some models, such as those used in urban transportation planning, still work with vehicle trips. This is probably because vehicles are the aggregate unit which is most directly related to the policy questions of interest, such as capacity, noise, pollution, etc.

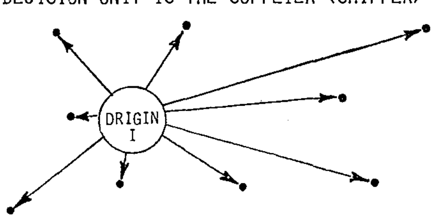
Working with tons is better. Individual commodities can be identified. Imbalances in flow by direction can be handled. Different types of equipment can be used and train lengths and barge flows are no longer a problem. There is, however, a fundamental problem with shipment size. The problem results from the fact that shipments of different sizes pay different transport charges even where all other aspects of the shipment are identical (i.e. commodity, origin and destination). The freight charges per unit can drop to one half, one third, one quarter or less as the size of the shipment rendered increases. Using tons as the basic element ignores this factor.

By working directly with a shipment as the basic element, the size problem can be addressed directly as one of the explicit choices to be made by the decision making unit. The interrelationships between the annual flow of a commodity by an establishment, the frequency of shipment and the shipment size should also be noted. If the annual flow rate is known, as it could be if the individual decision unit is identified, then the choice of shipment size will also result in knowing the frequency of shipment.

4) **Base the Model at the Destination** — If we are to model the behaviour of individual decision making units, then the unit to be addressed will be either the shipper or the receiver. It cannot be both, because to do so would involve double

ORIGIN BASED MODEL

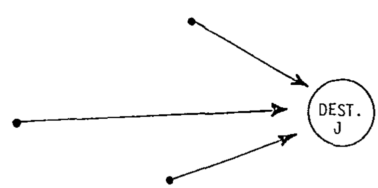
DECISION UNIT IS THE SUPPLIER (SHIPPER)



- SUPPLIERS TYPICALLY MAKE DECISIONS ON PRICE AND PRICING POLICY (F.O.B., C.I.F.)
- A SUPPLIER TENDS TO PRODUCE ONLY A FEW PRODUCTS
- SUPPLIERS SHIP OUTPUT TO MANY DIFFERENT CONSUMERS IN MANY DIFFERENT CITIES
- COMPETITION BETWEEN PRODUCERS IS DIFFICULT TO REPRESENT WITHOUT MANY TO MODEL
- DIFFICULT TO IDENTIFY DECISIONS AND ATTACH BEHAVIDRAL SIGNIFICANCE
- DIFFICULT TO SPECIFY AND ESTIMATE MODEL BECAUSE OF THE LARGE NUMBER OF DESTINATIONS, THE DIFFICULTY WITH DECISIONS, ETC.

DESTINATION BASED MODEL

DECISION UNIT IS THE CONSUMER (RECEIVER)



- CONSUMERS TYPICALLY MAKE DECISIONS ON CHOICE OF SUPPLIER, QUANTITY TO PURCHASE, FREQUENCY OF PURCHASE AND CHOICE OF MODE IN FOB PRICING (70% OF CASES)
- A CONSUMER OBTAINS VARIETY OF INPUTS FROM ONLY A FEW SUPPLIERS IN FEW LOCATIONS
- COMPETITION BETWEEN SUPPLIERS CAN BE EASILY INCORPORATED INTO A SINGLE MODEL
- DECISIONS ARE EASILY IDENTIFIED AND HAVE A CLEAR BEHAVRIORAL INTERPRETATION
- RELATIVELY EASIER TO SPECIFY AND ESTIMATE MODEL SINCE THERE ARE FEW SUPPLIERS AND DECISIONS ARE MORE EXPLICIT

CONCLUSION: DESTINATION BASED MODEL IS SUPERIOR

Figure 3 - Comparison of origin and destination based models

counting. We could presumably sample existing distributions of industry and population to get the individuals to be examined. There are a number of reasons why the receiver of commodities, rather than the shipper, is the appropriate decision making unit to sample in the general case. There are obviously cases in the real world where the shipper is actually making the decisions and it should be possible to modify the models to reflect this difference in point of view where it is appropriate. However, it is necessary to base the model either at the origin or the destination end of the shipment and our contention here is that the more appropriate place is the destination, for a variety of reasons. These reasons are summarized in Figure 3.

First, the decisions made by suppliers concern themselves primarily with price, and pricing policy. This may include whether the commodity is to be priced on an FOB factory price or CIF delivered price basis. By contrast, consumers typically make decisiosn of the choice of supplier, the quantity to purchase, the frequency of purchase, and the choice of mode in those cases where the price is quoted FOB. Since in some 70% of the cases prices are given FOB factory, the bulk of the travel/demand related decisions are made by the consumer rather than the supplier.

Another point of contrast has to do with the easier modelling approach which can be adopted on a destination based model. By and large, a supplier tends to produce only a few products, and to ship these products to many consumers in many different cities. A consumer, on the other hand, typically uses as input a large number of products, each from only a few suppliers in a few locations. Thus, if we work with a single commodity at a time, it is easier to work on the consumer side than the supply side, since the competition between suppliers can be relatively easily incorporated into a simple choice model at the destination end, whereas the competition between producers is much more difficult to represent without a many-to-many model.

Another argument for the destination-based model is that travel-related decisions are more easily identified, and have a clearer behavioral interpretation, whereas, on the origin side it is much more difficult to identify the decisions and to attach behavioral significance. For example, the decisions identified above on the part of the consumer as to choice of supplier, quantity and frequency of purchase, and choice of mode are understandable choices faced by the decision maker given the price of the product at each origin and the transport level of service available to the destination. On the supply side, the decisions are harder to interpret, however. There is no reason, for example, why the supplier does not ship to every destination, except that he has not been asked to do so by the consumer. Thus, in the final analysis, the consumer makes the decisions about where his supplies will be obtained.

This all adds up to a situation in which it is ultimately easier to specify and estimate a destination-based model, since there are fewer suppliers to deal with in the choice, and the decisions are more explicit. By working with the full distribution of industries, including agriculture, mining, manufacturing, wholesale and retail trade services, and those final-demand elements, such as population, government, investment, inventories, and exports, all receiving elements can be covered.

Our conclusion is that a destination-based model is easier to develop and easier to use than an origin-based model, and there are no corresponding disadvantages to working at the destination end.

5) **Determine Model Parameters Empirically** — The advantages of an empirically determined model are clear. One does not have to guess at the values which a shipper places on time, on loss or damage in transit, or on any of the other attributes associated with transport of the shipment. There is some question, however, as to how this can be done. Value of time, for example, cannot be observed directly—they must be inferred from choices made by the receiver. To do this, a situation must be identified in which choices involving different tradeoffs between cost and time can be observed. This is, in fact, not difficult to do, since each shipment made in the real world is proof of some choice. There is no dirth of choices. There is a problem, however, in determining the choice set actually faced by the shipper. The dependent variable in this case is either a zero or a one. Only that item chosen out of the set of possible choices would receive a one; the remainder are indicated by a zero. The relative frequency of a given choice could be developed for those choices with similar attributes. This suggests that a probability model might be useful. Consumer choice theory could then be used to infer values of the unknown parameters in the utility function.

Since each shipment involves a choice on the part of the decision maker, transportation waybills are the paper transaction proof of this choice. A waybill, or other shipping document, contains all the pertinent information. It typically has the commodity moving, its origin and destination, the mode by which it is travelling, the shipment size, and the number of pieces. From the name of the consignor and consignee, the industry and firm size of the establishments involved can be inferred. Ordinarily, it is possible to obtain the freight charges, and from time to time, even the transit time from a waybill. Also, from the using firm's industry and size, it is even possible to infer the annual use rate of a given commodity, though this would require some economic sleuthing. By obtaining samples of the waybills for given shipments at the destination-end, the receiver of the goods is well known, and pertinent information can be developed concerning him. Also, the sample of input commodities should be quite robust, since most firms use far more items as input than they produce as output.

There is, therefore, an obvious empirical base for the determination of freight-demand models—literally millions of waybills are produced daily. If these waybills could be obtained and used in the estimation procedures, an empirically-based model could be developed. These same procedures will also work for shipments by private carriers (i.e., private truck, pipe-line, barge), since in almost every case, there is some shipping document. The key question is the ability to capture this document and to supplement it with other information concerning the establishment making the shipment decisions.

6) **Formulate the Model to Use Generalized Attributes** — The use of generalized attributes to describe the commodities, the modal service offerings, the market or suppliers and the receivers is crucial to producing an economical and usable model. There is no area in which this is more true than with the different commodities. If it were necessary to formulate and calibrate a separate model for each commodity, the utility of the entire effort is in question. If, on the other hand, one generalized model can be built for a wide range of different commodities, with the individual commodities described by their attributes, then a much more workable approach is possible. Even if it becomes necessary to segregate into broad classes of commodities, such as bulk goads, particulates, liquids, packaged goods, etc., the number of classes can be workable. Commodity attributes for practically any level of commodity detail are available. The MIT Freight Transport Research Group has prepared a commodity attribute file for commodities at a 5-digit STCC level [2]. This commodity attribute file contains information on the value per pound, density, shelf life, state, and environmental protection requirements for approximately 1200 commodities. Thus, the use of generalized commodity attributes allows the development of a single set of commodity abstract freight demand models that can subsequently be used for a wide range of commodities, even those never before observed.

Generalized attributes can also be used to describe the level of service variables for the modal service offerings, as mentioned before. The models developed may then be either generic or mode specific, but since there are only a few modes in comparison to the large number of commodities, it is possible to estimate mode specific models if this proves to be more desirable.

The use of generalized supplier and receiver attributes allows the models to be specified generally, so that they may be used for a wide range of industries and firm sizes. Same care must be taken to keep the attributes general, rather than industry-specific, so that a generalized set of models can result.

7) **Base the Aggregation Scheme on Forecastable Data Sources** — If a disaggregate behavioral model is used, then it will ultimately be necessary to develop an aggregation scheme for use in forecasting. The forecasting scheme does not actually have to be developed at the time the model is developed, since the aggregation scheme is more directly related to the policy questions being addressed. If, for example, a national level forecast of mode choice is desired, it may be unnecessary to develop the flows state-by-state. Instead, a representative sample of receivers can be developed with observations from a wide range of geographical regions, industry types and commodities, and the model results for the sample can be aggregated directly and expanded to produce the national estimates. If, on the other hand, state-to-state estimates were required, then the sample used would have to include sufficient observations to be able to develop the additional detail needed state-by-state.

It is possible to work with small homogenous groups rather than a representative disaggregate sample. If small groups are used, the group means for each of the groups can be used in the forecast and the results summed over all groups to produce the overall forecast.

Regardless of which scheme of aggregation is used for the U.S., the future population industry and firm-size distribution use in forecasting can be developed from the very large data sets provided by the Census Bureau, by the BEA, and by the Bureau of Labor Statistics. The BEA regions in particular have forecasts to several future years already prepared by the issuing agency. These forecasts could be used in developing the sample used in the model.

Another possible source of observations are the Dunn and Bradstreet files. These files contain information on more than three million commercial establishments in the United States. There are also more than 300,000 Canadian establishments recorded. More than 367,000 establishments are reported for the state of California alone, covering all of the industry sectors—agriculture, forestry and fisheries, mining, contract construction, manufacturing, transportation, communication and public utilities, wholetrade, retail trade, finance, insurance, and real estate and services. These files are continuously updated to record only the existing population.

Both of these basic sources—the Census data and Dunn and Bradstreet—can be used to develop a sample of establishments in any area of the country. They could serve, therefore, as the forecastable data source upon which future freight demands can be developed.

CURRENT STATUS OF FREIGHT DEMAND MODELLING

At present, the principle barrier to the development of a set of policy-sensitive, disaggregate behavioral demand models for freight is the existence of an appropriate disaggregate data set for use in estimating such a set of models. The U.S. Department of Transportation has recently entered into a contract with the MIT Center for Transportation Studies to gather the required data and to develop a set of models for the United States. Explorations of the status of existing data are currently underway, and preliminary model specification has already been performed. Current thinking as to the specification of these models is included as Appendix B to this paper. The models described are formulated with simultaneous choice of point of origin, choice of shipment size and mode. They embody all of the desirable attributes described above.

Work on the models will proceed in three phases: Phase I includes model specification and pilot data collection. Phase 2 involves full-scale data collection, and Phase 3 involves model calibration and testing.

The models developed will be disaggregate at the level of the individual decision maker. They will, therefore, be extremely flexible. They could be used either individually to predict the choices made by an individual establishment for the transport of an individual commodity, or they could be used in aggregated form either embodying a disaggregate random sample or some more elaborate aggregation scheme. It is anticipated that the results will surpass those of other models developed to date. The literature review described in Appendix A tends to support this contention.

APPENDIX A

Literature Review

Relatively little work has been done in the area of freight demand modeling in comparison to the extensive body of literature on modelling the demand for passenger transportation. Nevertheless, a fairly large number of studies of freight demand have appeared in the transportation and economics literature during the past ten years.

In reviewing freight demand models, primary consideration should be placed on the policy sensitivity and completeness of each model. One measure, of a model's policy sensitivity is the extent to which it includes transportation level of service variables which are under the control of carriers and regulators. As described earlier, the list of level of service variables includes rate, mean travel time, and travel time reliability.

The second criteria is completeness. One aspect of completeness relates to the range of decisions addressed by a model. Models which predict only the choice of mode are less complete and less useful in policy analysis than models which cover the mode, shipment size and O-D choices. Another aspect of completeness relates to the range of situations in which a model can be applied. Some models can be used to forecast flows only for the commodities represented in the estimation data set, while other models can be applied to any commodity. Also, some models can be used to study the demand in only one region, while other models are transferable to any region. Hence, the completeness criteria is a measure of the applicability of a model to a wide range of demand related freight transportation problems.

Table A. 1 - Summary of Freight Demand Models

Type of Model	Modeler	Level of Service Variables	Commodity Descriptors	Other Variables	Functional Form	Principal Data Sources
Aggregate-Volume by Mode	Sloss, 58	rate		economic activity measure	log linear	19
	Perle, 51	rate	dummy variables, (stratification)	regional dummy variables	log linear	1,4
	Miller, 43	change in rate		production index	linear	1
	Black, 12	distance	(stratification)	regional consumption and production	gravity	2a, 2c
	Morton, 44	rate	(stratification)	production indices	log linear	5,7
	Wang, 60	rate		production indices	log linear	7,8
	Tihansky			GNP, production indices, modal shares	linear log linear	7
Aggregate-Systems	Mathematica, 39 Volume 2	rate, travel time	value, size distribution	GRP, population, sales, employment, area	linear	2, 9, 10
	A. D. little, 1	distance, circulty index	value, bulk dummy, seasonality dummy	population, employment, production and consumption factors	linear, special	1, 2b, 6
	Swerdloff, 61	distance	(stratification)	population, employment	share model	1, 2, 6
	Krosge, 36	rate, travel time, time variability, L&D	cost factor for LOS variables	consumption, investment, exports, production	linear program, min. cost	11
Aggregate-Modal Shares	Herendeen, 29	rate, travel time	(stratification)		log linear	1, 2a
	Perle, 51	rate	dummy variables, (stratification)	regional dummy variables	log linear	1,4
	Miller, 43	rate, stratification by distance	stratification by size	percent of firms with rail siding	linear	1, 2a
	Surti, 60	distance	size, (stratification)		linear	2
	Boeing, 13	rate		GNP	log linear	12
	Mathematica, 39 Vol. 3, p. 24	rate, travel time	value, (stratification)	annual volume	(special)	2, 20
	Mathematica, 39 Vol. 3, p. 30	rate, travel time	value, (stratification)	annual volume	(special)	2, 20
	Kullman	rate, travel time reliability, distance	value	annual volume	logit	1, 2, 13
	Roberts, 54	rate, travel time reliability, L&D	value, perishability packaging cost	stockout cost, variability of usage, cost of capital	linear	—
	Mathematica, 39 Vol. 1	rate, travel time, time reliability	value	frequency of orders variability of usage	(special)	—
	American Airlines, 3	rate, distance, pickup charge	value, density packaging cost	inventory, safety stock, cost of capital	linear	—
	Miklius, 42	distance	size, (stratification)	employment at the origin firm	linear discriminator	2
Disaggregate-Mode Choice	Miklius, 42	distance	size, (stratification)		linear	2
	Antle, Haynes, 5	distance, travel time, rate, handling cost	size, (stratification)	annual volume of shipments	linear discriminator	14
	Army, 30	distance, travel time, rate, handling cost	size, (stratification)	annual volume of shipments	linear discriminator	16
	Beuthe, 11	rates, fixed costs	size, (stratification)			15, 20
	Hartwig, Linton, 28	rate, travel time	value, (stratification)		logit, probit, linear discriminator	17
	Ruijgrok, 56	rate, travel time	dummy variables for state	outgoing dummy, industry variables	logit	18

Note: The number after the modeler's name refer to the bibliography

Note: Due to space limitations, it is impossible to list all co-authors, variables and data sources.

Table A.1
Key to Data Sources

1. *Carload Waybill Statistics.*
2. *Census of Transportation.*
   a. Volume 3, Part 1 Shipper Groups
   b. Volume 3, Part 1 Geographic Areas
   c. Volume 3, Part 1 Commodity Groups
3. *Freight Commodity Statistics, Motor Carriers of Property.*
4. *Freight Commodity Statistics, Class I Railroads.*
5. *Waterborne Commerce of the United States,* Part 5.
6. *Transportation Facts and Trends.*
7. *Survey of Current Business.*
8. *County and City Data Book.*
9. *Federal Reserve Statistical Release.*
10. Reports from the Columbian Ministry of Transport.
11. Civil Aeronautics Board Form 41.
12. *Census of Manufacturers.*
13. Survey of 63 Firms in the Ohio River Valley.
14. Reports from the Chicago Board of Trade on Grain Shipments.
15. Survey of 97 Firms in the Arkansas River Valley.
16. A Sample of 1213 Waybills from a Midwestern Shipper.
17. Mail Survey of Shippers made by Dutch Ministry of Transport.
18. Reports from Canadian Dominion Bureau of Statistics, and Principal Counterparts.
19. Carrier's Tariffs.

Demand models can be separated into two general groups: aggregate and disaggregate. Furthermore, the aggregate and disaggregate models can be grouped according to their dependent variable. A summary of previous freight demand modelling studies is given in Table A.1.

Aggregate Models of Intercity Freight Demand

Most of the freight demand studies done to date have utilized aggregate data from government sources. However, empirical work with the *Census of Transportation* and other similar data sets has brought to light several serious problems arising from the use of aggregate data.

The results of most aggregate freight demand studies have been disappointing. In particular, those models which encompass several choices are reportedly more difficult to estimate than single choice models (such as mode split models). However, this does not imply that single choice models are superior. These results simply imply that better data is required for the estimation of a complete system of models.

Aggregate Mode Choice Models

One of the best known studies of freight demand was conducted by Perle (1964). Perle postulted a model of mode split between common carrier truck and rail as a function of the rates. The data used in this study came from the *Carload Waybill Statistics — State to State Summary* and the ICC *Motor Carrier Freight Commodity Statistics.* The data were aggregated into five commodity groups: products of agriculture, animals, mining, forestry, and manufacturing. The data were also aggregated into the nine geographic regions used by the ICC in reporting the truck data. A time series of five years of this type of data was prepared.

The model split used as the dependent variable in Perle's model was computed on the basis of tons of shipments. The explanatory variables, rates were computed as total revenue divided by total tons. Perle's model is of the following form:

$$\log(V_{m1}/V_{n2}) = \beta_0 + \beta_1 \log(r_{m1}/r_{m2}) + \sum_{i=1}^{9} c_i R_i + \sum_{i=1}^{5} d_i y_k + \sum_{k=1}^{5} f_k c_k$$

where

V_{m1} = volume carried by truck
V_{m2} = volume carried by rail
r_{m1} = average revenue/ton on truck shipments
r_{m2} = average revenue/ton on rail shipments
R_i = (1 for region i, 0 otherwise)
Y_i = (1 for year i, 0 otherwise)
C_k = (1 for commodity k, 0 otherwise)

Perle estimates this model using ordinary least squared regression. The commodity dummy variables were found to be the most powerful explanatory variables. The regional variables had some impact, but the time variables were all insignificant. Perle concluded that the explanatory power of the rate term was minimal.

In an effort to improve the fit of this model, Perle stratified the data by commodity, by region and by both region and commodity. Models were then estimated on each subset of the data using the appropriate dummy variables in each case. The results of his work were very mixed. Some models fit very well, while others had large residuals and insignificant coefficients. Estimates of the price elasticities varied widely depending on the level of aggregation. In general, the effects of the commodity and region dummy variables were more significant than the effect of the rate term.

The results reported by Perle are not surprising. The dummy variables used for commodities are correlated with many of the important commodity attributes and transport level of service attributes. In particular, the commodity variables acted as a proxy for value per pound. And since value is correlated with rates, the commodity variables are correlated with rates. Furthermore, the regional dummy variables acted as a proxy for travel time reliability, loss and damage, and other level of service variables which vary significantly between regions (especially for rail transport).

Several conclusions can be drawn from Perle's work. First even simple mode split models require a more complete set of commodity and level of service variables. Secondly, the problem of aggregation bias in the values of the coefficients can be quite severe. Thirdly, aggregate level of service variables are neither good explanatory variables, nor good policy variables. The rate variable turned out to be very weak in all of perle's models. And in terms of policy analysis, the average revenue per ton is too vague to be of much use because it includes such a wide range of commodities and lengths of haul. Thus is can be concluded from Perle's study that the use of more level of service attributes, more commodity attributes, and more disaggregate data is desirable.

The conclusions drawn from Perle's study are reinforced by a study conducted by Miller (1972). Miller proposed a model of the rail market share as a function of the rates and a measure of rail availability. The rail market share was computed for each weight-mileage block in each of the 85 shipper classes included in the 1967 *Census of Transportation.* An average rail rate corresponding to each weight-mileage block in each shipper group was computed from a special tabulation of the 1965 *Carload Waybill Statistics*. No suitable source of truck rates could be located and therefore the truck rate variable was dropped from the model. Rail availabilitiy was measured as the percentage of plants with rail sidings, using data from the 1967 *Census of Manufacturers*.

The general form of Miller's model is the following:

$$(V_{ml} / {}_{m} V_{m}) = \beta_0 + \beta_1 (r_{ml}) + \beta_2 \text{ (rail availability)}$$

where

V_{ml} = volume carried by rail

r_{ml} = average rate on rail shipments

A separate model was estimated for each weight-mileage block. In general, the results were poor. In most cases the availability term had a significant coefficient, but the rate variable did not. Miller tried aggregating the data over weight blocks and estimating a model using only the rate variable. As expected, the rate variable had a significant coefficient in this second version of the model. However, when the availability variable was put back into the model and a third estimation was attempted, the rate variable was again insignificant.

These results are not surprising. The influence of rail rates on modal shares is largely a function of the rates on the competing modes. Thus the lack of a truck rate variable in this model makes the rail rate variable difficult to interpret. It should also be noted that rail availability is one outcome of the plant location decision. The plant location decision is influenced by the transport level of service attributes, even though this strategic choice is not very sensitive to short-run fluctuations in the level of service. Therefore, the rail availability variable captured part of the influence of travel time, reliabilty, loss and damage, as well as the rates. The problems with the model could have been mitigated by using these level of service variables explicitly in the model. It is also evident that a greater disaggregation of data is needed to allow a more precise definition of the level of service variables (including rates) which influence demand in particular market segments.

Another study of modal split was conducted by Surti and Ebrahimi (1972). These researchers estimated a model of truck-rail mode split using the data on the tons of shipments in each weight-mileage block of the 24 shipper groups in the 1963 *Census of Transportation*. A separate model was estimated for each shipper group. The length of haul was used as a proxy for the level of service variables and shipment size was used as a proxy for other logistics costs. The data on both of these independent variables were also taken from the Census.

The most successful version of their model is of the following form:

$$V^k_{m1} / (V^k_{m1} + V^k_{m2}) = \beta_0 + \beta_1 (\text{dist}) + \beta_2 (q)$$

where

V^k_{m1} = volume of commodity k carried by truck

V^k_{m2} = volume of commodity k carried by rail

q = shipment size

This model fits most shipper groups fairly well. All estimated coefficients have significant t statistics and all r^2 statistics are above 0.80. Note that these results are better than one might expect based on the experience of Miller (1972). The reason for this is a subtle difference in the specifications of these two models. Because of his stratification scheme, Miller actually estimated a model of mode choice conditional on shipment size and distance, but not commodity type. Since Miller's model lacked commodity attributes, the variation in commodities undermined his results. In contrast, Surti and Ebrahimi stratified their data so that their model represents the mode split conditional on the type of commodity. Therefore the lack of commodity attributes in the Sutri/Ebrahimi model caused no major problems. Furthermore, since the mode and shipment size choices are made jointly, shipment size should be a good explanatory variable of mode choice. However, the usefulness of the Surti/Ebrahimi model is limited because of the lack of level of service variables. Rates and travel times are policy sensitive, but distance is not.

A somewhat wider variety of variables was included in a rail-barge mode split study conducted by A. D. Little Inc. (1974). The data for this study came from the 1967 *Census of Transportation*, the 1966 *Waterborne Commerce of the United States*, and the 1966 *Carload Waybill Statistics*. The variables used in this model are:

V^k_{ij} = volume of commodity k shipped from i to j

v = value/ton of commodity k

d = distance from origin i to destination j by rail

c = circuity index = (water distance/rail distance)

S = (1 for seasonal goods, 0 otherwise)

B = (1 for bulk goods, 0 otherwise)

L = percentage of production facilities located on the water at the origin plus the percentage of consuming facilities located on the water at the destination.

Note that the variable L is similar to the availability measure used in Miller's study. Also, distances are used as a proxy for rates as in the Surti and Ebrahimi study. However, this study includes some different variables as well. Three commodity attributes (v, S, and B) are used, in addition to a market attribute (V_{ij}).

The functional form of the A. D. Little model is the following:

$$\sin^{-1}\sqrt{V^k_{ij,m1} / (V^k_{ij,m1} + V^k_{ij,m2})}$$
$$= \beta_0 + \beta_1 \log(V^k_{ij}) + \beta_2 \log(v) + \beta_3 \log(d) + \beta_4 \log(L) + \beta_5 \log(C) + \beta_6 (B) + \beta_7 (S)$$

where

$V^k_{ij,m1}$ = volume of k carried from i to j by barge and

$V^k_{ij,m2}$ = volume of k carried from i to j by rail

This model was estimated for each of five geographic regions. Within each region, modal shares were computed for flows between BEA zones of 17 commodity groups (including raw materials and finished products).

The results from estimating this model were mixed. The r^2 statistic varied from 0.2 to 0.64. All of the coefficients had the expected sign and most were significant, except for the coefficient of the variable L. Note that the problem with the variable L is similar to the problem with the availability variable in Miller's model. Both studies indicate that the correlation between long run decisions such as plant location and various level of service and commodity variables is strong enough to force some key variables to have insignificant coefficients. However, this does not imply that plant location should be excluded from mode split models when level of service attributes and commodity attributes are used. Often the long run decisions are sub-optimal with respect to the current situation. Uunder these circumstances, the correlation between the long run decision variables and the level of service attributes will be lower, and terms like L will tend to add a significant amount of explanatory power to the model.

Several researchers have attempted to specify aggregate mode split models in which the mechanism for decision making is somewhat more apparent in the model structure. One such model was proposed by the consulting firm Mathematica (1969). The model that was proposed is the following:

$$V^k_{ij,m1} / (V^k_{ij,m1} + V^k_{ij,m2}) = 1 / [1 + (AVC_{m2} / AVC_{m1})^{\beta_1}]$$

The important feature of this model is that the variable AVC has been defined in the following manner:

$$AVC_m = rate_m + \beta_2 (time_m * value) * \beta_3/[V^k_{ij}]^{0.5}$$

The first term of this expression represents the out-of-pocket transport cost and the second term represents the in-transit carrying cost. The third term is designed to reflect the inventory carrying cost. Together these three terms add up to an approximation of the average variable cost of using mode m to transport commodity k from origin i to destination j. The advantage of this kind of specification is that it incorporates a comparison of the logistics cost of the shipment alternatives. It should be noted that this model addresses freight demand at a more disaggregate level than the models previously discussed. This allows variables such as rates, transit time and commodity value to be more precisely defined.

The Mathematica model was estimated for each of 15 commodity groups using data from the 1963 *Census of Transportation* on rail, truck, and air shipments. Rates were estimated for all three modes using models developed for this study. Crude procedures for estimating travel times for each mode were also developed. In general the estimation results were good. Most coefficients in the set of estimated models were significant and many of the r statistics were above 0.80. These encouraging results tend to support the opinion that this Mathematica model was a step in the right direction.

In the same paper which was discussed in the previous review, Mathematica (1969) proposed another model. This second model does not make use of logistics cost variables. Instead, ratios of the level of service variables are used to compare the two competing modes. The form of this model is the following:

$$V^k_{ij,m1} / (V^k_{ij,m1} + V^k_{ij,m2}) = 1 / (1 + w)$$

where:

$$w = [(t_{m1} / t_{m2})^{bc_{m1}} (c_{m1} / c_{m2})^{bt_{m1}} (c_{m1}^{\ln(t_{m1})} / c_{m2}^{\ln(t_{m2})})^{b}]^{u}$$

u = $[b*\ln(v) + b*v]\,[b*\ln(V^k_{ij}) + b*(V^k_{ij})]$

t_m = mean travel time from i to j by mode m

c_m = tariff on mode m for shipment of k from i to j

v = value of commodity k

V^k_{ijm} = volume of commodity k sent from i to j by mode m

This model performed about as well as the other Mathematica model. But this second model suffers from the drawback that its parameters are much harder to interpret than the parameters of the first model.

Kullman (1974) also tried to develop a mode split model with a clear interpretation. Kullman assumed that the cost of shipping by a given mode could be expressed as a linear function of the level of service attributes, commodity attributes and market attributes. The independent variables used in this model include high-way distance, annual tonnage, commodity value, rates, mean travel times and a measure of the variation in travel times. These variables were used in a logit form model of the rail-truck mode split.

$$\log (V^k_{m1} / V^k_{m2}) = \beta_0 + \Sigma_i \beta_i x_i$$

where x_i is an explanatory variable, and

V^k_{m1} = volume of commodity k carried by rail

V^k_{m2} = volume of commodity k carried by truck

Unlike the first Mathematica model, the independent variables used by Kullman are not estimates of logistics costs. He simply substituted rates, travel times and other independent variables for the x's used in the formula shown above. Kullman experimented with three sets of flow data which came from the 1967 *Census of Transportation.* The first includes national level mode splits for 2, 3, 4 and 5 digit commodities. The second data set contains mode splits for 2, 3, 4 and 5 digit commodities which were shipped between Production Areas and Market Areas. The third data set is a special preparation of the Census data. It includes mode splits on flows between counties of high, medium and low value goods.

The empirical results from Kullman's study were disappointing. The r^2 statistics were low and there were many insignificant coefficients in the models that were estimated. One conclusion that can be drawn from this study is that data without geographic detail *and* commodity detail *and* market/firm detail is not adequate. This study reinforces the conclusion that a model which is sensitive to the full set of level of service variables must be estimated with disaggregate data.

Aggregate Systems of Models

Several attempts have been made to build systems of aggregate models which are capable of covering the full range of freight shipment decisions. Typically these systems consist of a series of one decision models organized along the lines of the Urban Transportation Model System. A sequential model system is acceptable if it includes feedback from short run decisions to long-run decisions. However, this has not been adequately modeled in the systems which have been developed to date.

Systems of aggregate models suffer from the same problems that plague individual aggregate models. They may not be transferable in space or time because the estimates of the coefficients depend (in an unknown way) on how the data has been aggregated. Also, systems of aggregate models may not contain some policy variables because the aggregation of data tends to reduce the explanatory power of key variables such as travel time reliability. Nevertheless, the demand modeling systems currently available do offer a simple methodology for doing comprehensive freight planning.

The A.D. Little mode split model discussed earlier in this chapter has been used as part of a system of models developed by this firm (A.D. Little, 1976). The mode split model was reviewed separately because it has several particularly interesting features. The other elements in this system of models will not be reviewed here, although they are referred to in the summary table.

One aggregate model system of interest was developed by the consulting firm Mathematica (1969) as part of the Northeast Corridor Transportation Project. This system is composed of four stages. The first stage involves a projection of the total production in each of 16 commodity groups. The projections are made with a separate regression equation for each group. The independent variables in these regressions include a time variable and projections of various segments of the GNP. The GNP projections must be provided from an outside source.

The second stage involves a projection of the regional share of originating and terminating tonnage in each commodity group. In the final version of the model, it was assumed that the regional shares of originating tonnage remain unchanged. The regional demand for each commodity is predicted using a regression model. The independent variables in this model include population, retail sales, per capita income and regional income. Projections of these independent variables must be provided from other sources.

In the third stage, a distribution model is used to predict interregional flows. An initial guess is provided by a regression model which uses the following independent variables: production at the origin, consumption at the destination, distance, and various socio-economic variables such as population and employment at the destination. But when flows are predicted in this manner, the total flow in and out of each region will not match the totals predicted in the second stage. Therefore, a flow adjustment algorithm was developed using Lagrange multipliers. The objective of the Lagrangian is to minimize the flow of adjustments subject to the constraints on the total flow in and out of each region.

The final stage in the system involves the modal split of the inter-regional flows. A separate market share regression model was used for rail, common carrier truck, private truck, air, water and "other". The independent variables used in these models include the fraction of shipments falling into each of five weight groups, the fraction of shipments falling into each of eight distance groups, commodity value and average gross revenue per ton. Note that when these mode split models are used, the shares must be normalized so that they total to 100 percent.

Mathematica's system of models was calibrated with data from the 1963 *Census of Transportation*. The data base included flows in 16 shipper groups between 25 Production and Market Areas. Supporting data came from the *City and County Data Book* (Bureau of the Census), "Business Statistics" (Dept. of Commerce), and the "Federal Reserve Statistical Release". Unfortunately, information on the performances of the complete system was not included in the report.

Another system of sequential aggregate models has been developed by the Office of Systems Analysis (1970) in the Department of Transportation. The data base for this study was built around a 506 zone system that covers the entire country. Networks connecting these zones were constructed for rail, truck, water, air refined product pipelines and crude pipelines. In this model system, flows are classified as being petroleum or non-petroleum. Non-petroleum flows are subdivided into large and small shipments. Both large and small shipments are further divided into three value classes. Petroleum products are divided into crude and refined.

The first step in this study was to build base year inter-zonal flow tables for each commodity group. Air flows were estimated using CAB data on the commodity flows in and out of all major airports. A gravity model was used for flow distribution. Barge flows came from a special preparation of *Waterborne Commerce*. Pipeline flows were estimated by applying a linear programming model to data on the production and consumption of crude and refined petroleum in various zones. Truck flows were estimated from an inter-county motor vehicle trip table prepared from data collected by the Bureau of Public Roads. In preparing the truck flows, auto trips were "factored out" of vehicle trips and then average truck load factors were applied to the remaining highway volumes.

Projections of inter-zonal flows are made using the Fratar model which was developed as part of the Urban Transportation Model System. The Fratar model has been used to adjust interzonal flows so that they will be consistent with the zonal in-flows and out-flows projected in the previous step. The independent variables in this model are the changes in zonal population and employment.

Adjustments in model split are made using a share model of the following form:

$$V^{k}_{ij,m1} / {}_{m}V^{k}_{ijm} = (\beta_{1,m1}\, t_{m1}^{\beta_{2,m1}}\, r_{m1}^{\beta_{3,m1}}) / ({}_{m}\, \beta_{1,m}\, t_m^{\beta_{2,m}}\, r_m^{\beta_{3,m}})$$

where

t_m = mean travel time from i to j by mode m

c_m = tariff on shipments from i to j by mode m

The time and rate variables used in this study were derived from the minimum path distances in each of the model networks. Regression equations relating distance to rate were estimated using I.C.C. data on the costs and revenues of each mode.

This model system has been tested with a number of policy scenarios. The results were reportedly reasonable. However, the details on the system have not been widely publicized.

Aggregate, Joint Demand Models

Single choice models can be assembled into sequential model systems which address the full range of freight shipment decisions. However, there are two drawbacks to this approach. The first is that some choices (such as mode choice) are made jointly with other choices (such as shipment size). Secondly, even when two decisions are not made jointly, there is feedback from short-run decisions to long-run decisions. Neither of these two aspects of freight demand are adequately represented in sequential model systems.

The problems with sequential model systems have given rise to joint or direct, aggregate demand models. The advantage of this approach is that several choices are modeled in the same equation. In theory, the independent variables can be structured in such a way as to reflect the combined effect of a set of decisions. The independent variables could represent the interactions between choices and the model coefficients would then reflect the importance of various interactions. In practice, this approach has not been used to its full advantage. Most applications of aggregate joint demand models have involved a combination of the trip generation and mode split elements of the sequential model systems. However, the level of production and mode of shipment are usually not chosen jointly. This makes it difficult to specify independent variables which reflect the interaction of these two choices. Consequently, most aggregate, joint demand models have been constructed around two separate sets of variables: the mode choice variables and the volume of production variables. In this respect, these models are more like two separate models contained in the same equation. Whatever interaction effects are represented in the model, they are imbedded in the coefficients.

A joint aggregate demand model was estimated as part of Perle's (1964) study which was described earlier. The data set used to estimate this model is the same as the one described before. It includes truck and rail flows in five commodity groups, in nine regions, during each of five years.

The model used by Perle is of the following form:

$$\log(V_{m1}) = \beta_0 + \beta_1 \log(r_{m1}) + \beta_2 \log(r_{m2}) + \sum_{i=1}^{9} c_i R_i + \sum_{i=1}^{5} d_i Y_i + \sum_{k=1}^{5} f_k C_k$$

where

V_{m1} = volume of traffic carried by mode m1
r_{m1} = average revenue/ton on mode m1
r_{m2} = average revenue/ton on mode m2
R_i = (1 for region i, 0 otherwise)
Y_i = (1 for year i, 0 otherwise)
C_k = (1 for commodity k, 0 otherwise)

Perle estimated a truck model and a rail model of this form. In general his results were very poor. In all cases, the results from this model had poorer r^2 and t statistics than Perle's aggregate mode split model.

These results are to be expected. The dependent variable in the joint model includes the choice of a level of production as well as the choice of a mode. In contrast, the Perle model previously described covers only the choice of mode. Obviously the joint model taxes the explanatory power of the data more heavily than the mode split model. However this does not entirely explain the difference in results.

The most crucial flaw in the joint model is that it does not reflect the fact that the demand for transportation is derived from the demand for commodities. The dependent variable includes the volume of transportation, but none of the independent variables explain the demand for the commodities being transported. It is true that the price of transportation is a component in the sales price of a good, which in turn determines the demand for that good. However if this rationale is to be used, then the appropriate variable to put in the model is the sum of the cost of transportation and all other costs associated with the production of a good. But where all commodities are aggregated into a small number of groups, the average cost of production for each group is almost meaningless. On the other hand, it is impractical to estimate a separate demand model for each commodity. As will be shown, other researchers have found methods of using proxy variables to represent the demand for commodities. Nevertheless, Perle's study does reinforce the conclusion that aggregate models are inherently difficult to specify properly.

Another important study in this area was conducted by Sloss (1971). Sloss postulated a model for the volume of truck traffic as a function of the average truck rate, the average rail rate and a proxy variable used to represent the demand for commodities.

One unique aspect of this work is that Canadian rather than U.S. data were used. The dependent variable was defined as the annual tons of freight carried in intra-provincial, interprovincial and international hauls by trucks registered in each province. The sources of information on this variable are the "Motor Transport Traffic: National Estimates" published by the Dominion Bureau of Statistics, and the provincial counterparts of this report. These same reports were used to collect data on the average revenue per ton for truck hauls, which were used to estimate average truck rates. The average rail rates were measured in terms of the average revenue per ton for intraregional FCL shipments of selected commodities. Data on this variable came from the "Waybill Analysis" published by the Canadian Board of Transport Commissioners.

Unlike Perle, Sloss used a measure of economic activity in his model to represent the demand for commodities. This variable was defined as the sum of farm cash income, the value of new building permits and the value of shipments of manufactured goods in each province. Data on this variable came from the "Canadian Statistical Review" and the *Canada Yearbook*.

Data were collected for eight provinces for the years 1958 through 1963. Then ordinary least squares was used to fit the following model:

$$\log(V_{m1}) = \beta_0 + \beta_1 \log(r_{m1}) + \beta_2 \log(r_{m2}) + \beta_3 \log(E)$$

where

V_{m1} = volume of truck traffic

r_{m1} = average revenue/ton on truck

r_{m2} = average revenue per ton on rail

E = economic activity variable

The results of Sloss' work indicate demand elasticities of nearly unity with respect to each of the three independent variables. Although the r^2 statistic was quite high, the estimation results are not conclusive. The reason for this skepticism is that the data used in this study was so highly aggregated that almost all variability was lost. This implies that very different results might be reported if this model was estimated using data on much smaller geographic units. Unfortunately, this is a problem which plagues all aggregate models to some degree.

Alexander Morton (1969) has conducted a demand modeling study using data similar to Perle's and the same model specification as Sloss. The data on rail volumes was taken from "Freight Commodity Statistics for Class 1 Railroads" which is published by the ICC. The 242 commodities listed in this report were aggregated into five groups: products of agriculture, animals, forestry, mining and manufactures. Truck volumes were taken from the American Trucking Association pamphlet titled "Transportation Facts and Trends". Using data from the same source, truck rates where calculated as total revenue divided by total ton-miles. Rail rates were calculated from the RI-1 index of relative rates, which was published as part of the I.C.C. "Rail Waybill Study". The data were gathered for the years 1947 through 1966, for the nation as a whole and selected regions. The economic activity variable used in this study was GNP for the nation and gross regional product for regions.

Morton estimated the model for truck and rail, using various subsets of the data. He also estimated a similar model in which the truck and rail rates were replaced by the average rate on both modes, and the ratio of truck and rail rates. The results of this work varied considerably with the level of geographic and commodity aggregation. Due to the aggregation of data, the r^2 statistics were fairly high, ranging from 0.58 to 0.94. However, over one-quarter of all coefficients estimated in this study had the wrong sign. Morton attributed part of the problem to the historical shift from rail to truck caused by level of service factors other than rates. This demonstrates once again how the exclusion of key variables can undermine a model.

Disaggregate Models

For purposes of policy analysis, a demand model must be able to forecast aggregate patterns of freight movements. In theory, this can be accomplished by aggregating the data on the independent variables before they are used in the model, or by using disaggregate data in the model and then aggregating the results. It was shown in the preceding section that the aggregation of the data on the independent variables has led to major problems in many studies. These problems can be avoided if the model is estimated using disaggregate data.

The advantages of disaggregate models are numerous. One of the most important points is their efficient use of data. Since the data is not averaged, there is no loss in the variability (i.e., explanatory power) of the independent variables. This means that reliable estimates of the model coefficients can be obtained from relatively small data sets. Furthermore, disaggregate models often contain significant coefficients for variables that usually have insignificant coefficients in aggregate models. This is particularly true of policy sensitive variables such as travel time reliability.

A second important feature of disaggregate models is that they are potentially transferable. This means that an estimated disaggregate model which is properly specified can be applied to a wide range of commodities and markets.

Another feature of this kind of model is that forecasts can be prepared for any level of aggregation. Hence it is not necessary to have separate sets of models for local, regional and national planning.

One point that should be emphasized is that disaggregate models require data on the atributes of all of the available freight shipment options, both the chosen and unchosen. Although the collection of this kind of data may seem like a nuisance, it does allow the modeler to view the shipment process from the point of view of the decision maker. All of which means that the independent variables can be defined clearly and concisely, and the coefficients can be interpreted unambiguously. Furthermore, any a priori knowledge of the manner in which decision makers evaluate alternatives can be incorporated into the specification of the model.

Because of the lack of data, very few disaggregate freight demand studies have been conducted. To date, there have been no attempts to estimate a joint choice model, although several mode choice models have been estimated.

A disaggregate mode choice model was estimated by Antle and Havens (1971) at the Institute for Water Resources. The independent variables used in this study are the following:

x_1 = shipper's annual volume of shipments of given commodity between given O-D pair

x_2 = length of haul

x_3 = average travel time

x_4 = average shipment size

x_5 = rate on chosen mode

x_6 = difference in rates between chosen and alternative mode

x_7 = handling cost on the selected mode

This data was collected for coal, coke, and petroleum shipments in the Ohio River Valley. The dependent variable was defined as having the value 1 if barge was chosen and 0 if rail was chosen.

The modeling technique used in this study is known as discriminant analysis. The form of the model is the following:

$$Z = \sum_{i-1}^{7} \beta_i x_i$$

When using the model, if the computed value of X exceeds a critical value, then the model predicts that barge will be chosen, otherwise the model predicts that rail will be chosen.

The results from the estimation of this model are fairly good. All of the coefficients came out with the expected sign and most were significant, although the distance, annual volume and rate variables were weaker than expected.

Antle and Haynes also tried aggregating their data across all commodities and then re-estimating the model. The results were significantly poorer. This supports the claim made earlier that disaggregate models use data more efficiently than aggregate models.

The latest attempt at estimating a disaggregate mode split model is described in a thesis written by Hartwig and Linton (1974). These two researchers collected 1213 waybills from one shipper of consumer durables. Using the data from the waybills, they calculated the rate, mean travel time and variance in travel time for the full truckload and full rail carload alternatives. Commodity value was also included as an independent variable.

Hartwig and Linton used this data to estimate logit, probit and discriminant analysis models. Although the logit and probit models performed quite well, the travel time variable was insignificant in most of the specifications which were estimated. Nevertheless, this study is important because it provides further evidence of the practicality of estimating disaggregate freight demand models.

The first attempts at estimating disaggregate freight demand models are encouraging. However, the problem of building a joint choice model and including a wider range of independent variables has yet to be tackled.

APPENDIX B

Formulation of Disaggregate Freight Demand Model [3]

This section develops a general specification for a disaggregate freight demand model. This model can then be specialized for the case of the urban/regional planner or for the case of the national transportation policy planner.

The impetus for a disaggregate approach to freight demand modelling originated in the urban passenger transport field, where recent breakthroughs in the modelling of individual travel behaviour have occurred. Urban transportation researchers, after years of frustration with aggregate models of travel behavior, concluded that the choice of transport mode is intertwined with choices of workplace, residential location, and auto ownership, as well as household characteristics such as family size, income, and number of workers. The response to this condition was to develop models (or model systems) that would make, for the individual household, probability estimates of choosing each of the possible combinations of mode to work, workplace, residential location, etc., based on household characteristics and transport levels of service.

In the freight field, a similar situation exists. The shipper must make three simultaneous choices: where to buy, how much to ship, and by what mode (or carrier), based on annual requirements for a commodity, storage, ordering and other costs, the levels of service offered by competing modes, the price of the commodity quoted at different origins, and characteristics of the commodity (shelf life, value, packaging, special handling requirements, etc.). As in the passenger transport field, it is conceptually appropriate to formulate a choice model which assigns probabilities to combinations of shipper alternatives (origin, shipment size, and mode) based on shipper, transport mode and commodity attributes. [4]

The mathematical form most frequently specified for disaggregate choice models are multinomial logit functions of the form:

$$P(\underline{X}|\underline{A}) = \frac{e^{U(\underline{A})}}{\sum e^{U(\underline{A})}}$$

$\underline{X}$ = vector of choice combinations

$\underline{A}$ = vector of attributes

P = probability of choosing a particular combination $\underline{X}$*

U = utility function based on all the attributes

For the case of freight demand prediction, this general form specializes as follows:

$$P^k(i,mq|ALTS) = \frac{e^{U(T,C,M,R)}}{\sum e^{U(T,C,M,R)}}$$

where

U(T,C,M,R)	=	the utility function of the receiver
k	=	commodity index
i	=	supply (origin) point
mq	=	mode/shipment size combination
ALTS	=	alternatives avaible to the receiver
U	=	utility function
T	=	transport attributes
C	=	commodity attributes
M	=	market attributes
R	=	receiver attributes

Figure 1 defines the T, C, M and R variables that could enter the utility function.

Transport Attributes
W = wait time (days)
T = transit time (days)
R = reliability (days)
L = loss and damage (unitless, $0 \leq L \leq 1$)
\$ = freightrate (\$/1b)
C = special charges (\$/1b)

Market Attributes
P = relative price (unitless)
O = ownership (binary 0-1)

Commodity Attributes
V = value (\$/1b)
D = density (1b/ft^3)
S = shelf life (days)

Receiver Attributes
A = annual use rate (1bs/year)
M = mixed order (unitless, $0 \leq M \leq 1$)
S^1 = seasonal purchase (unitless, $0 \leq S^1 \leq 1$)
Q = shipment size (1bs)
U = reliability of use rate (days)
G = guarantee of availability (unitless %)

Figure 1 - Variables that can Enter a Utility Function

The task of the freight transportation analyst is quite clearly the *specification* and *estimation* of the utility function U(T,C,M,R). While several specifications for U(T,C,M,R) are possible, they are all estimated using maximum likelihood techniques.

Specifying the utility function of the receiver can be done in light of the logistics process he/she is trying to manage. Basically, total costs consist of purchase costs plus logistics costs, as follows [5].

Total Costs = purchase cost + order and handling cost + transport cost + capital carrying and storage cost + stockout cost

The utility function of the shipper is developed by combining the variables previously specified with appropriate parameters. See Figure 2. By careful specification, the parameters can be presented in such a way that they can be interpreted as constants, interest rates, elasticities or dimensionless. This would allow the estimated models to be checked for reasonableness and extended to other environments where estimation is not practical. It would even allow the model to be used without estimation should that be necessary.

The utility function in Figure 2 is based on the "classic" calculation of logistics costs, with parameters taking on values that represent various aspects of a shipper's/receiver's cost structure. The equation is constructed to yield a disutility measure in units of \$/lb., i.e., the *total* cost of the commodity from time of purchase until time of consumption. Each cost element in Figure 2 is discussed below:

Purchase Cost — P, the relative price at each origin, is multiplied by V, the value per pound delivered, to obtain a local price. Both P and α_1 are dimensionless, and α_1 should be equal to unity.

U (T,C,M,R,) = purchase cost + order and handling cost + transport cost + capital carrying and storage cost + stockout cost

Purchase Cost

cost to buyer

$\alpha_1 P \cdot V$

Order & Handling Cost

set up charge

$\alpha_2 \, 1/Q \cdot M$

Transport Cost

capital cost in transit — loss & damage — perishability — transport charges

$$\alpha_3 \frac{W+T}{365} \cdot P \cdot V \cdot 0 + \alpha_4 (L \cdot P \cdot V)\, 0 + \alpha_5 \left(\frac{W+T}{S}\right)^n \cdot P \cdot V \cdot 0 + \alpha_6 (\$+C)\, 0 \cdot M/Q$$

Capital Carrying Cost & Storage Cost

$$\alpha_7 \left(\frac{Q}{2A} + \frac{R+U}{365}\right) \cdot V \cdot P \cdot S^1 + \alpha_8 \left(\frac{Q}{2A} + \frac{R+U}{365}\right) \cdot S^1/D$$

Stockout Cost

stockout

$\alpha_9 (1-G) \cdot (1/Q)$

Figure 2 - Specification of the Utility Function

Order and Handling Cost — the cost of the personnel and paperwork required to order each shipment, where 1/Q = F/A. The term M varies from 0 to 1.0 and is the percent by weight of this commodity in a mixed shipment. Mixed shipments, in which more than one commodity is involved but where the sum of all commodities is less than a full truckload, are transported under the freight rate applicable to be highest rated commodity applied to the combined shipment weight. Both transport costs (see α_6 below) and order costs must be apportioned over all items in the shipment. The term α_2 will be the cost of placing each order.

Transport Cost — the first term represents the capital carrying costs of the goods while in transit; 0 is a 0-1 variable which signifies ownership. If the receiver is buying the goods f.o.b. the origin, 0 takes on the value 1. If the receiver is buying f.o.b. the destination, the shipper bears these charges and 0 = 0. The parameter α_3 is simply the cost of capital for the receiver.

The loss and damage term hinges on the transport attribute L which is derived from a separate model and represents the fraction of units totally destroyed. The cost of damaged units is subsumed in L. The parameter α_3 should calibrate to unity.

The perishability term has the units of \$/lb., and α_5 is consequently expected to calibrate to unity. The parameter n on the term $[(W+T)/S]^n$ is designed to modify the influence of this term and must be specified exogenous to the model. For example, if the commodity is fresh fruit, n may lie in the range $0 \leq n \leq 1$ to reflect the fact that as W+T approaches S, there will be a significant los due to spoilage. Alternatively, if the commodity does have a finite shelf life but does not lose value until W+T is very close to S, then n will take on a positive value greater than 1. The determination of a proper value for n will be left to analyst judgment and the experience of the traffic managers who assist in data preparation.

Transport charges are the freight rate per pound times the binary variables 0 to indicate who pays the freight and the mixed shipment variable M described above. Dividing by Q distributes the freight charges on a per pound basis. The parameter α_6 should calibrate to unity.

Capital Carrying & Storage Cost — the first term is the cost of the merchandise while in the receiver's warehouse prior to consumption. The term Q/2 is the average level of stock on hand (exclusive of safety stock) and the term (R+U)/365 represents the safety stock required to protect from transit time unreliability and usage rate unreliability. The variable R is the number of days beyond the mean transit time in which there is a probability G that the shipment will arrive, given a constant rate of daily commodity use. The variable U represents the variation in the rate of use of stock and is measured by the standard deviation of the (presumed normally distributed) use rate. The term S^1 represents the influence of seasonality and is the fraction of the year that the item is held in inventory. Hence, $0 \leq S^1 \leq 1$. Note that for an item which is special ordered as needed (i.e., is not held in inventory), $S^1 = 0$ as would be expected. For an item used for a production run which lasts only four months, $S^1 = .33$. The parameter α_7 should calibrate to the cost of capital.

The second term is an expression for the costs of warehousing and represents the average amount of goods on hand $[Q/2A + (R+U)/365] \cdot S^1$ times the reciprocal of the density. This expression is valid for either bulk or packaged commodities. The term α_8 will be the cost of storage per *cubic* foot. (Note that warehouse costs are normally calculated on a per square foot basis so that the stacking height of cartons becomes key. However, it is too difficult to generalize about this variable, so it was not put in the model). Thus, the parameter α_8 for packaged commodities will pick up the influence of both warehouse costs per square foot and stacking height which will increase the variance of this parameter estimate.

Stockout Costs — the final term is a representation of the cost per pound of stocking out, which occurs with probability 1-G each time there is a shipment (F times per year) and is distributed over the annual use A, where F/A = 1/Q. The parameter α_9 is a measure of the cost of each incidence of stockout. The variable G is measured in terms of the probability that the shipment arrives in time to prevent a stockout and will typically range from .90 to 1.0.

This brief discussion of logistics costs is in no way intended to substitute for a careful reading of any of the available texts on physical distribution management. It is intended only to provide an overview of the way that logistics terms might enter the model. In a given situation the regional planner would likely have to add, delete, or modify terms to meet local data requirements.

The four attribute vectors (T,C,M,R) can be developed in the following way. *Receiver attributes* are determined through exercise of the industry firm size table and input/output model coefficients

$$A^{K}_{IND,SIZE} = a_{K,IND} \cdot \hat{X}_{IND,SIZE}$$

where:

$A^{K}_{IND,SIZE}$ = annual usage rate of commodity K by firm of size class SIZE in industry IND

$a_{K,IND}$ = input-output coefficients

$\hat{X}_{IND,SIZE}$ = average output of all firms of industry IND and size class SIZE

Note that the annual usage rate developed by this procedure will only be as detailed as the input-output table used. It is possible that regional planners may have access to input-output tables that differ in degree of detail from national tables which are typically at the 3-4 digit SIC. Also, the result is in dollars. To convert to physical units the result must be divided by the value per pound, using the commodity attribute described below. Other receiver attributes, such as available facilities, whether the commodity is used as an intermediate or final good and whether the receiver uses mixed orders are a function of the receiving industry. Once the receiving industry is known, these inputs can be quickly determined.

Transport level of service variables can be developed for a given situation using three separate classes of level of service models by mode. These models, developed by the Freight Transportation Group at MIT are: [6]

1) Waiting and Transit Time and Reliability Models
2) Loss and Damage Models
3) Freight Rate Estimation Models

Each is described in detail in the references; however, a brief description of each is given below.

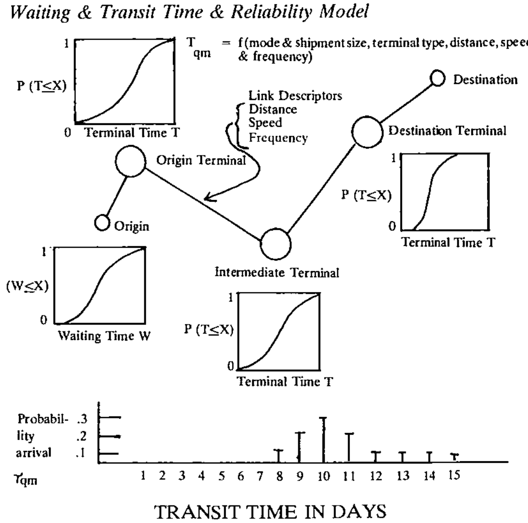
The waiting and transit time and reliability models predict time distributions for waiting plus transit from origin to destination as a function of number and type of terminals and the line haul distance, speed and frequency of service between terminals for each of the modes. See Figure 3. Since the principle cause of delay in the system is that which occurs at terminals this approach has produced very good comparisons with observed travel time disributions measured in the real world [7]. The probability of delay at a given type of terminal is represented by a cumulative function of the time available between arrival and the next regularly scheduled departure.

Loss and damage by mode is a function of the commodity attributes and the particular transport mode under consideration. The models used are simple regression models based on the experience record of the mode.

Freight rate estimation using a model is essential because of the complexity of the commodity rate structure and its huge size (more than 3 trillion separate commodity tariff rates are on file in the U.S.). The model uses as input the various commodity attributes (i.e., density, value per pound, shelf life, etc.) and the distance and shipment size by mode. The regression models have been developed using actual waybill or freight bill information for the various modes. Though point estimates of the freight rate are produced, it is possible, using the error distribution produced by the regression, to predict the distribution of likely freight rates if this becomes desirable.

The commodity attributes, the third class of variables, are available for 1200, 5-digit STCC commodities from the MIT Commodity Attribute File [8]. This file uses the Standard Transportation Commodity Classification (STCC) code at the five digit level to record the following information in machine readable form:

Waiting & Transit Time & Reliability Model



Loss and Damage Model

L^{k}_{mq} = f(Mode & Shipment Size, Commodity Attributes)

Freight Rate Estimation Model

R^{k}_{mq} = f(Transport Attributes, Commodity Attributes, Mode & Shipment Size, Distance)

Figure 3 - Features of the Level of Service Models

1) STCC Code No.
2) 35 digits of Description
3) Wholesale Value per pound (\$/lb)
4) Density (lbs/ft[3])
5) Shelf Life (weeks)
6) State (solid, liquid, gas, particulate)
7) Environmental protection required (frozen, temperature, pressure, shock)

Given the commodity, this information can be made quickly available.

Market attributes are more difficult to secure. The most important is price. The price variable is designed as a *relative price*, which when multiplied by the wholesale value per pound becomes a local price. Using relative price enables differential model prices to reflect the spatial distribution in prices in the input. Even wholesale and retail markups can be simulated. Price and availability information can be obtained from the Office of Business Economics, the Agricultural Marketing Service, and the Bureau of Mines of the U.S. Government. Or, this data may be furnished by a macro-economic model used in conjunction with the study. Data on ownership and facilities is normally a function of the industry from which the commodity of interest is drawn. For example, the food industry normally sells its products with CIF delivered prices and has rail sidings available for loading rail cars.

Model Outputs

The output of the model is the probability that a particular receiving firm located in the region will secure its input from origin i in shipment size q by mode m. When this probability is multiplied by the annual use rate, $A^k_{IND, SIZE}$ calculated from the industry/firm size analysis previously described, the result is the commodity k moving from each of the known producing regions to the mode choice/shipment size for each region for the firm under consideration.

$$V^k_{i, qm, IND, SIZE} = p^k(qm|i) \cdot A^k_{IND, SIZE} \tag{8}$$

where:

$p^k(qm|i)$ = probability of shipping commodity k by mode m at shipment size q from origin i

$A^k_{IND, SIZE}$ = annual use rate of commodity k by industry IND of size SIZE

When summed over all firms and firm sizes in the region, the results can be stored in a single three-dimensioned table. See Figure 4. If the commodity moving is of no particular interest, then the result should be presented using only two dimensions.

$$V_{i, qm} = \sum_k V^k_{i, qm} \tag{9}$$

A third possible approach is to summarize flows by major commodity grouping or segment. In this case:

$$V^{kseg}_{i, qm} = \sum_{k \varepsilon kseg} V^k_{i, qm} \tag{10}$$

kseg = aggregation of individual commodities k

This minimizes the extent of the third dimension to some reasonable size from the full size used for the analysis (k at 5 digits) would be approximately 1200).

Obviously, the computations to be performed are voluminous if the number of commodities is large and the number of origin regions is extensive. The process described above is, in fact, merely an enumeration over all decision makers. This number of computations can be reduced by sampling instead of using total enumeration. The best dimensions for sampling appear to be those concerning industry, IND, and firm size, SIZE. However, the commodity, k, also looks like it should be sampled, particularly because of its size. On the other hand, the dimension i, covering each origin region is a good choice for enumeration since there is considerable reason to preserve spatial detail.

A random sampling process can be set up for selecting the representative sample so that the commodity k to be used is selected based on the relative size of the Kth row of the input-output table. Finally, a firm size is selected based on the relative size of the appropriate row in the industry/firm size matrix. Each such randomly selected point is used for computation until a sufficiently large sample is available that the total, by mode and shipment size, from the different supply points can be adjusted up to equal the total volume flowing.

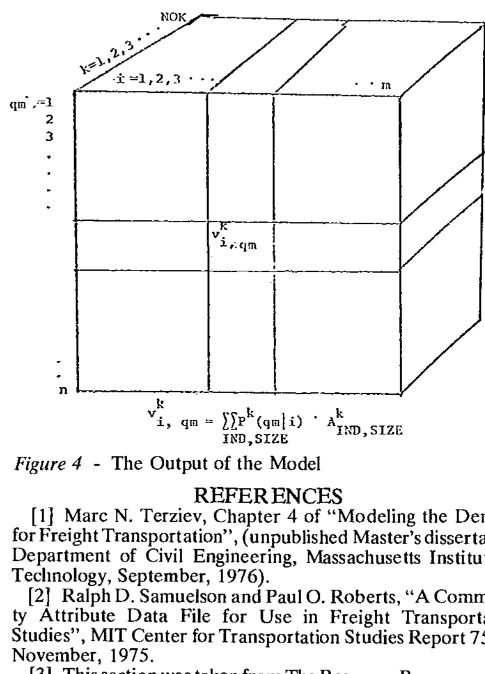


Figure 4 - The Output of the Model

REFERENCES

[1] Marc N. Terziev, Chapter 4 of "Modeling the Demand for Freight Transportation", (unpublished Master's dissertation, Department of Civil Engineering, Massachusetts Institute of Technology, September, 1976).

[2] Ralph D. Samuelson and Paul O. Roberts, "A Commodity Attribute Data File for Use in Freight Transportation Studies", MIT Center for Transportation Studies Report 75-20, November, 1975.

[3] This section was taken from The Resource Paper prepared by Paul O. Roberts and Brian C. Kullman for Presentation to the Workshop on Goods Movement at the Third International Conference on Behavioral travel Modelling, April, 1977, Adelaide, Australia.

[4] P. O. Roberts, "Forecasting Freight Flows Using a Disaggregate Freight Demand Model, MIT Center for Transportation Studies, CTS Report 76-1, January 1976.

[5] Roberts, Paul, **The Logistics Management Process as a Model of Freight Traffic Demand,** International Symposium on Freight Traffic Models, May 1971, Amsterdam, Holland.

[6] Roberts, Paul O., **Transport in Intercity Markets: An Overwiew of the Physical Distribution System,** Task II report, CTS no. 75-17, October, 1975.

[7] Martland, Carl D., **Improving Railroad Reliability: A Case Study of the Southern Railway, Volume 10, Appendix B,** MIT, 1975.

[8] Samuelson, Ralph and Paul O. Roberts, **A Commodity Attribute File for Use in Freight Transportation Studies,** CTS Report no. 75-20, November, 1975.