

MODELING THE DEMAND FOR FREIGHT TRANSPORTATION

by

MARC NICHOLAS TERZIEV

B.S.E., Princeton University  
(1974)

Submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science in Civil Engineering

at the  
Massachusetts Institute of Technology  
September, 1976

Signature redacted

Signature of Author . . . . .  
Department of Civil Engineering, June 14, 1976

Signature redacted

Certified by . . . . .  
Thesis Supervisor

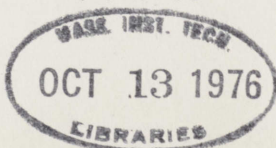
Signature redacted

and . . . . .  
Thesis Supervisor

Signature redacted

Accepted by . . . . .  
Chairman, Departmental Committee on Graduate Students  
of the Department of Civil Engineering

Archives



MB  
7-21-78

MODELING THE DEMAND FOR FREIGHT TRANSPORTATION

BY

WANG NICHOLAS TIKLIS

S.S.A., Princeton University

(1974)

*Thesis*

Submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Civil Engineering

*CR*  
*1976*  
*MS*

at the

Massachusetts Institute of Technology

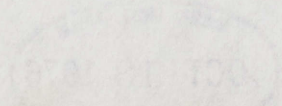
September, 1976

Signature of Author . . . . .  
Department of Civil Engineering, June 16, 1976

Certified by . . . . .  
Thesis Supervisor

and . . . . .  
Thesis Supervisor

Accepted by . . . . .  
Chairman, Departmental Committee on Graduate Studies  
of the Department of Civil Engineering



ABSTRACT

## MODELING THE DEMAND FOR FREIGHT TRANSPORTATION

by

MARC NICHOLAS TERZIEV

Submitted to the Department of Civil Engineering on June 14, 1976, in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering.

This study is concerned with the development of policy sensitive models of the demand for freight transportation. Models of this type are useful in the analysis of a wide range of transportation issues, including modal regulation, pollution, energy conservation and investment in the infrastructure.

This study begins with the development of a conceptual framework. The demand for freight transportation is determined by the way in which firms order supplies. Each order involves the choice of a supplier, a mode and a shipment size. These three choices are usually made jointly because each one affects the cost of the other two. The joint decision is motivated by a desire to minimize the sum of the purchase cost and logistics cost. The logistics cost includes the transport cost, capital carrying cost, stockout cost, and loss and damage cost. The magnitude of these costs will be affected by commodity, market and receiver attributes, as well as the transport level of service.

Once the key variables and relationships have been identified, the literature on freight demand modeling and the sources of data are reviewed. Most of the models developed to date have utilized aggregate data. This type of data has limited the usefulness of these models. Better policy analysis models could be developed with disaggregate data. However, very little disaggregate data are currently available. The principal shortcoming of the published data is the lack of a description of the types of firms which use each type of freight transportation. There is also a need for better data on commodity markets and the transport level of service.

In light of the conceptual framework and the shortcomings of existing models, a specification has been developed for a disaggregate model of the joint choice of a supplier, mode and shipment size. The theory of logistics management has been used to develop simple equations for each of the important cost factors.

Although the full implementation of the proposed model requires the collection of new data, preliminary empirical tests have been conducted using the available aggregate data. A model of the choice of mode for a given shipment size has been tested with several specifications. A model of the joint choice of mode and shipment size has also been tested. The results of these experiments indicate a need for more detailed data. Furthermore, additional research is needed for the development of disaggregate models which can handle the joint choice of discrete and continuous variables, such as mode and shipment size.

This study demonstrates that logistics management theory can be used to specify policy sensitive freight demand models. This appears to be a promising approach for future research.

Thesis Supervisor  
Title

Moshe E. Ben-Akiva  
Assistant Professor

Co-Supervisor  
Title

Paul O. Roberts  
Professor

ACKNOWLEDGEMENTS

I would like to thank my advisors, Professor Moshe Ben-Akiva and Professor Paul Roberts for their many insights and helpful comments.

I would also like to thank Kevin Tam for typing the first draft of this document.

Most of all, I would like to thank my friends and family for the encouragement they have given me during this research, and during the past six years. Special thanks should go to my sister Sandy, who helped type the final draft, and to Bob Menhard, who helped me keep my sense of humor throughout.

Table of Contents

	<u>Page</u>
Title Page . . . . .	1
Abstract . . . . .	2
Acknowledgements . . . . .	4
Table of Contents . . . . .	5
List of Tables . . . . .	8
List of Figures . . . . .	10
 Chapter 1 Introduction . . . . .	 11
The Supply-Demand Equilibrium in Freight Transportation	13
The Use of Freight Demand Models	17
Outline of the Following Chapters	18
 Chapter 2 Development of a Conceptual Framework . . . . .	 21
The Decision Making Process	21
The Hierarchy of Choices	25
Key Variables	27
The Logic Behind Freight Transportation Decisions	31
 Chapter 3 A Review of the Available Data . . . . .	 37
Sources of Intercity Flow Data	37
Sources of Market and Receiver Attributes	46
Sources of Level of Service Attributes	47
 Chapter 4 Literature Review . . . . .	 51
Aggregate Models of Intercity Freight Demand	57
Aggregate Mode Choice Models	59
Aggregate Systems of Models	75
Aggregate, Joint Demand Models	80
Disaggregate Models	88

## Table of Contents (cont.)

	<u>Page</u>
Chapter 5 A Breif Review of the Theory of Disaggragate Demand Models . . . . .	93
Attributes of Disagqregate Demand Models	93
The Structure of Disaggreagte Models	96
Data Considerations and Model Estimation	100
The Need for Further Research into the Formulation of Disagqregate Models	101
Chapter 6 Development of a Model Specification . . . . .	106
The Scope of the Model	106
Specification of the Independent Variables	110
Summary of the Specification	133
Aspects of the Decision Making Process Which Require Special Attention	134
Chapter 7 Preparation of a Data Base . . . . .	137
The Use of Commodity Flow Data	138
The Development of Receiver Attributes	145
The Development of Commodity Attributes and Level of Service Data	150
Summary	156
Chapter 8 Estimation Results . . . . .	157
The Conditional Mode Choice Model	157
Definition of the Set of Alternatives	159
The Naive Specification	160
Specifications Based on Logistics Costs Concepts	167
Summary of Results from the Estimation of the Conditional Mode Choice Model	178
A Model of the Joint Choice of a Mode and a Shipment Size	178
Summary	184

## Table of Contents (cont.)

	<u>Page</u>
Chapter 9 Conclusions and Recommendations for Further Research . . . . .	186
Directions for Further Research	189
Bibliography . . . . .	192
Appendix I Other Sources of Data . . . . .	198
Appendix II Production Areas and Market Areas . . . . .	203
Appendix III Shipment Sizes . . . . .	205



List of Tables

	<u>Page</u>
<u>Chapter 4</u>	
1. Summary of Freight Demand Models . . . . .	54
<u>Chapter 7</u>	
1. Origins and Destinations Included in the Data Base . . . . .	141
2. Cross Tabulation of Rail Shipments by Size and Distance . . . . .	143
3. Cross Tabulation of Truck Shipments by Size and Distance . . . . .	144
4. Freight Rate Estimation Models . . . . .	153
5. Mean Travel Times for Truck Shipments . . . . .	155
<u>Chapter 8</u>	
1. Definition of Variables Used in the Naive Specification . . . . .	162
2. Specification of the Cost Functions for the Naive Conditional Mode Choice Model . . . . .	163
3. Estimation Results from the Naive Model . . . . .	166
4. Definition of Variables Used in Specifications Based on Logistics Costs . . . . .	170

## List of Tables (cont.)

	<u>Page</u>
5. A Logistics Cost Based Specification of the Conditional Mode Choice Model . . . . .	171
6. Estimation Results from the Logistics Cost Based Specification . . . . .	174
7. Re-specification of the Logistics Cost Based Mode Choice Model . . . . .	176
8. Estimation Results from the Revised Logistics Cost Based Specification . . . . .	177
9. Shipment Alternatives Used in the Mode and Shipment Size Choice Model . . . . .	179
10. Definition of Variables Used in the Specification of the Joint Choice Model . . . . .	180
11. Estimation Results from the Joint Choice Model . . . . .	183

List of Figures

	<u>Page</u>
<u>Chapter_1</u>	
1. The Equilibrium Process in Freight Transportation . . .	15
2. The Equilibrium Process with Lag Effects . . . . .	16
<u>Chapter_2</u>	
1. The Hierarchy of Choices . . . . .	26
2. Key Variables in Freight Demand . . . . .	30
<u>Chapter_6</u>	
1. The Reordering Process When a Trigger Point System is Employed . . . . .	119
2. Probability Distribution of the Procurement Time for an Input . . . . .	120
3. Probability Distribution of the Daily Usage of an Input . . . . .	122
4. The Relationship Between the Reorder Point, the Safety Stock and the Probability of Stocking Out . . .	126
5. The Relationship Between the Average Size of a Stockout and the Probability of Stocking Out . . . . .	127

## Chapter 1

### Introduction

Determining the volume of freight traffic which will flow under a given set of circumstances is the starting point for any quantitative analysis of freight transportation policy. This is true whether the issues being studied are those concerning a carrier's level of service offerings, or a government's regulatory policies, or government investment in the transportation network. In spite of what appears to be an obvious need for analysis tools, very little has been done to provide such a capability in the freight area.

One may account for this situation in several ways. First, there is the lack of a comprehensive theoretical framework. This framework should include three main elements. They are: a conceptual model of the process by which the demand for freight transportation is determined; a mathematical model that can be used to test the conceptual model; and a method of using the mathematical model to analyze policies of interest. One of the primary goals of this thesis is to provide this kind of framework.

A second reason why better analysis tools have not been developed is the scarcity of detailed data. The lack of a

theoretical framework has caused a great deal of confusion as to exactly what kind of data should be collected. Thus, of the mountains of freight data which are available, only a small fraction is of any practical use in analyzing the policy issues that are currently of interest. Furthermore, much of the model development work has been motivated entirely from the standpoint of making use of the existing data. Therefore, many of the existing models are only marginally useful.

A third reason why the present stock of tools is inadequate is the increasing complexity of the policy issues in freight transportation. At one time the government's role was confined primarily to the regulation of rates and the construction of roads, canals and airports. Today government policies also address issues such as energy consumption, pollution, the rate of technical innovation, and the overall quality of service. In addition, the government must take a stand on the faltering health of many of the rail and air carriers. Clearly the analysis of these issues requires the consideration of a broad range of factors. Although an in-depth analysis of particular policies is beyond the scope of this thesis, the modeling methodology presented herein is sufficiently flexible that it could be used to forecast the impact of policies in many complex subject areas.

## The Supply - Demand Equilibrium in Freight Transportation

Before delving into the development of a theoretical framework in Chapter 2, it is helpful to review the supply and demand structure of the market for transportation services. The decision-makers on the supply side are the carriers. Given the volume of shipments being made, the carriers must decide on what level of service to offer [1]. Over the long run they can choose the quantity and type of vehicles, as well as the quantity and size of terminal facilities. In the short run, the carriers can adjust the frequency and reliability of the service. They may also adjust the rates, although only with the approval of the government regulators.

On the demand side there are the shippers and receivers. Given the level of service offered by the carriers, the shippers and receivers must decide on the quantity to send and the method of shipment. In the long run they can change the location of their businesses in response to changes in the supply of transportation. In the short run the shippers and receivers can choose the mode and shipment size that they use. Their decisions will be influenced not only by the market for transportation, but also the market for the commodity being shipped [2].

- 
1. The carriers' decisions are also influenced by government policies and regulations.
  2. The decisions of shippers and receivers are also influenced by taxes, labor markets and the availability of inputs.

The key elements in the equilibrium process are shown in Figure 1. The market system acts to bring the offerings of the carriers into equilibrium with the demands of the shippers and receivers. The equilibrium solution is characterized by the volume of shipments sent and the level of service actually experienced by these shipments.

In practice, the market for transportation services is never in a state of long run equilibrium. This is the case because the decision-makers do not react instantaneously to each others actions. One reason for this is that a finite amount of time is required to gather the information used in the decision making process. A second reason is that costly decisions are only re-examined when the situation is significantly altered. Even then, new decisions may not be implemented immediately because of their expense. Thus, the market system involves lags as indicated in Figure 2. When the lags are taken into account, it can be seen that the equilibrium process is actually a continuous cycle of actions and reactions. This study focuses on the behavior on the demand side; the responses of shippers and receivers to the level of service offerings of the carriers.

Figure 1

The Equilibrium Process in Freight Transportation

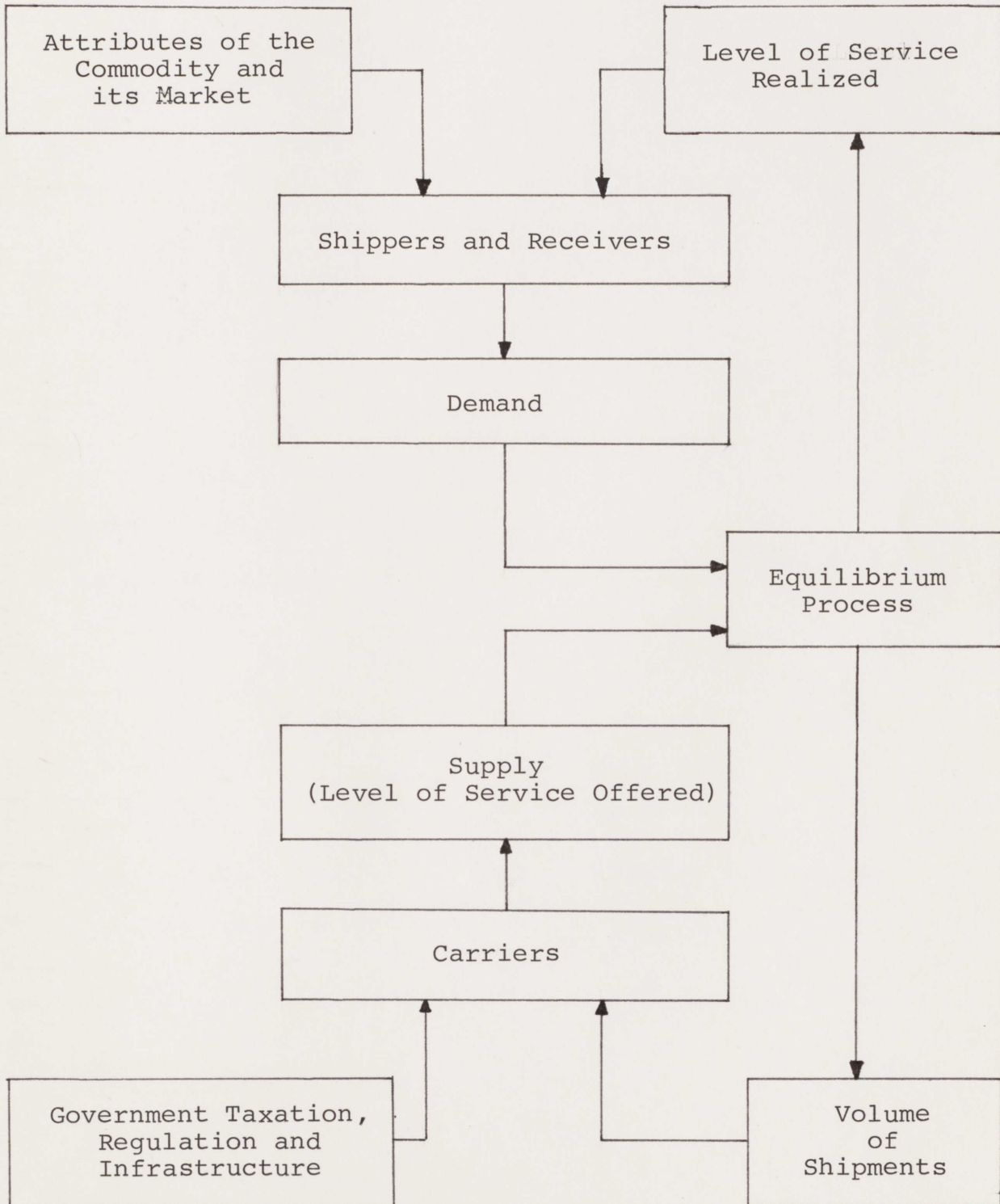
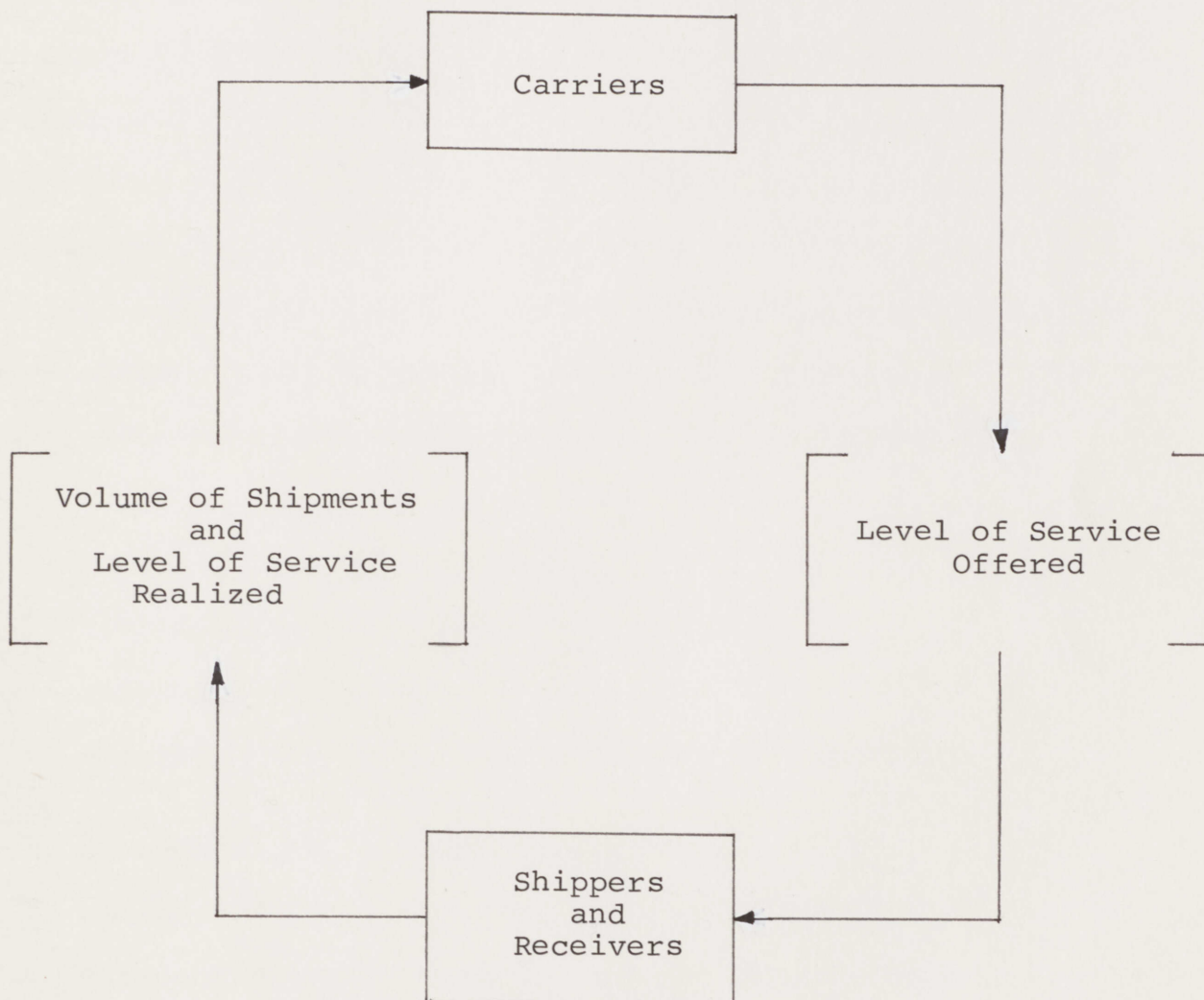




Figure 2

The Equilibrium Process with Lag Effects

## The Use of Freight Demand Models

A model that could forecast the behavior of shippers and receivers would be of use to government planners, policy-makers, and regulators as well as the carriers. The increasing economic problems of some freight carriers (the railroads and airlines in particular) has focused national attention on issues concerning the restructuring of the entire transportation system to aid the distressed carriers. Moreover, the need to economize on fuel and cut air pollution has led to controversial suggestions that the government should discourage the use of some modes. Some of the key issues of particular interest are the following:

- Deregulation of rail rates.
- Rationalization of the rail network.
- Advanced TOFC/COFC services.
- Improved internodel coordination and tariffs.
  
- Deregulation of truck rates.
- Easing of entry restrictions into trucking.
- Changes in truck size and weight regulations.
  
- Expansion of unregulated pickup and delivery services for air carriers.
- Changes in fuel cost and availability.
  
- Waterway user taxes.
- Continued federal sponsorship of improvements in the waterway system.

It should be noted that policies in all of these areas have one point in common. No matter whether a policy originates with the government or the carriers, its effect will be perceived by the shippers and receivers as a change in the level of service. Thus, the key to constructing a freight demand model that is useful in policy analysis is to include a wide range of level of service attributes. However, this step alone is not enough to guarantee the success of the model. The level of service attributes interact in a complex manner with the attributes of the commodity being shipped, and the attributes of the decision-maker. It is important that a policy sensitive freight demand model capture these interactions. One of the primary goals of this study is to build a methodological framework for accomplishing this.

#### Outline of the Following Chapters

The general philosophy introduced in this chapter is pursued in more detail in Chapter 2. A conceptual model of the decision making process on the demand side is developed from the point of view of an individual user of freight transportation services. Following this, a list of influential variables is developed. The theory of logistics management is used to gain insights into the nature of the decision making process and the role of the key variables.

Sources of freight transportation data are reviewed in Chapter 3. Special emphasis is placed on determining the availability of data on the key variables discussed in Chapter 2. The limitations of the data go a long way toward explaining the historical development of freight demand models. The stock of available models is reviewed and critiqued in Chapter 4. The primary purpose of this review is to determine the extent to which existing models have incorporated the important variables and relationships.

Chapters 5 and 6 are concerned with the development of a mathematical model that can be used to implement the conceptual model presented in Chapter 2. Specifically, Chapter 5 includes a discussion of the state-of-the-art in disaggregate qualitative choice models. These models have been applied very successfully in urban passenger transportation studies in recent years. However, the type of disaggregate model needed in the freight area is slightly different than those which are currently available. Some promising research on the development of a new disaggregate model is also presented in Chapter 5. Then in Chapter 6 the specification of the independent variables in the model is discussed. These variables are designed to reflect the important conceptual relationships, while still being practical from a data collection standpoint. Together, Chapters 5 and 6 define a model in specific enough terms to guide future data collection efforts. But, the model remains general enough that it

could be tailored to a wide range of situations.

Chapters 7 and 8 describe some model estimation experiments which have been conducted. Chapter 7 includes a discussion of the possibilities and pitfalls of using the existing data to estimate disaggregate models. The data base described in Chapter 7 has been used to test several model specifications. The estimation results are presented and analyzed in Chapter 8. Finally, conclusions and recommendations for further research are presented in Chapter 9.

## Chapter 2

### Development of a Conceptual Framework

The demand for freight transportation is derived from the demand for commodities in markets which are geographically removed from the locations at which commodities are produced. Thus, an analysis of the demand for freight transportation cannot be divorced from an analysis of the functioning of the market system. However, at an aggregate level the market system behaves in such a complex manner that the influence of any single factor is virtually unidentifiable. It is only at the level of the individual decision-maker that the interaction of transportation and other market factors can be studied in detail. In this chapter, the structure of choices made by a decision-maker is examined. The theory of logistics management is used to explain how these choices are related to the transport of goods, and how these choices are influenced by a variety of factors, including the behavior of the market system.

### The Decision Making Process

In freight transportation, the decision-maker is a manager of a manufacturing plant, a wholesale distributorship, or a retail store. It is the responsibility of the manager to set or

anticipate the daily level of output, and to assure that an adequate supply of inputs are on hand. These inputs are usually stockpiled according to an inventory control plan which is designed to offer some specified degree of protection against stockout. As materials are used from the stockpile, orders for the various inputs are generated in a manner specified by the inventory plan. Each order for a particular commodity involves the choice of a supplier (i.e. Shipment origin), a shipment size and a mode of carriage. In some cases the choice of mode is made by the supplier located at the shipment origin. But it can be argued persuasively that the supplier must act in the best interest of the customer if he wants to continue to do business with that party. For modeling purposes, we can assume that there is only a single decision-maker, who is located at the destination end of the shipment.

Although a manager has some alternatives available each time an order is placed, there are many suppliers, modes and shipment sizes which are not available in the short run. The reason for this is that less flexible long run decisions make it economically unattractive to ever use some modes, suppliers and shipment sizes. The longest run management decision is that of plant location. This decision is made with a general knowledge of the suppliers and markets in the region, and the quality of transport available. However, the location decision is (usually) not predicated on the choice of a particular supplier, mode and

shipment size. These three decisions may be altered from time to time, but the plant location decision will only be re-evaluated when there are major changes in regional markets and transportation services. Meanwhile, the plant location may give a commanding advantage to some subset of suppliers and carriers.

The choice of plant size (i.e. long run average level of output) is also a long range decision. This decision will put an upper and lower bound on the volume of inputs that will be required, which in turn will put some broad bounds on the set of feasible shipment sizes. This may preclude the use of certain modes which specialize in very large or very small shipments. Furthermore, the plant size decision may eliminate from consideration some suppliers who are not able to fill orders at a rate compatible with the volume of production.

Once the plant location and size have been chosen, the range of alternative suppliers can be narrowed down. Using rough estimates of the cost of transportation, a list of the most competitive suppliers can be made. In many cases the purchaser will have to enter into a multi-order contract with a supplier. Thus the choice of a supplier is in some respects more of an intermediate run decision than either the mode or shipment size choices. Nevertheless, the choice of a supplier is very closely related with the choices of mode and shipment size in both the short and long run.

Given the average level of production in the plant and a



description of the supplier(s), an inventory control strategy can be derived. This strategy will be designed to give whatever level of protection against stockouts that is deemed desirable by the manager. The choice of a risk of stockout will roughly define the range of feasible shipment sizes and the minimum required size of the warehouse used for stockpiling inputs. This decision will also lead to the development of guidelines for the minimum acceptable reliability of the transport mode. Thus, the list of feasible modes and shipment sizes will be shortened even further. However the exact values of the parameters of the control system will depend on the exact level of production in the plant and the choice of a particular combination of supplier, shipment size and mode.

Although the range of alternatives available in the short run is often limited by long run and medium run choices, there are usually a fairly large number of options left open. Within the range of feasible shipment sizes, there will probably be several competitive modes. There may also be some flexibility in choosing a supplier. It is important to note that the final decisions must be made jointly. The choice of a shipment size will heavily influence the transport level of service. Conversely, the choice of a mode will have an effect on the desirability of different shipment sizes. And the choice of a supplier will have a bearing on the relative attractiveness of both modes and shipment sizes. It is the responsibility of the

manager to take these relationships into account when he places an order for supplies.

### The Hierarchy of Choices

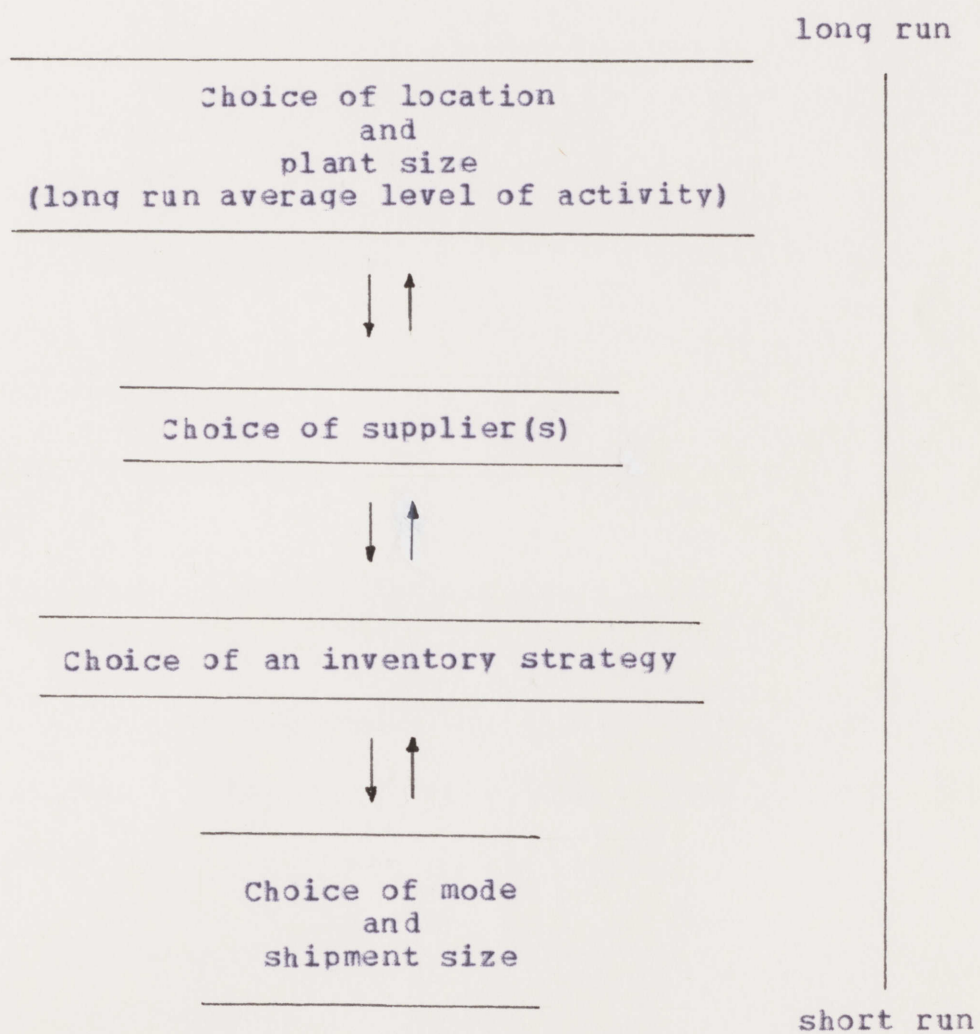
The discussion in the preceding section implies that the demand for freight transportation is determined by a complex hierarchy of choices. This hierarchy is depicted in Figure 1. The sequence of decisions that is assumed in this hierarchy reflects the different time lags involved in changing decisions in response to changes in the transportation system or the market situation.

It is important to note that this hierarchy does not imply one-way causality. There is feedback from short run decisions to long run decisions. Changes in the location of suppliers will eventually affect the plant location, just as changes in the chosen shipment size will influence the type of inventory control used. In the long run the causality runs in both directions.

The value of hypothesizing a choice hierarchy is that it gives some idea what the scope of a demand model needs to be to analyze policies in a particular time frame. For example, a model of the mode and shipment size choices might be suitable for an analysis of the immediate impacts of a small rail rate hike. But, the choice hierarchy indicates that the same model would probably be inappropriate for an analysis of the ten-year impact of a major rail rehabilitation policy.

Figure 1

## The Hierarchy of Choices



## Key Variables

There are four basic types of variables which affect the transport decisions described in the preceding section. One important influence is the level of service offered by each mode for various commodities, shipment sizes, origins and destinations. The key variables relating to the level of service are:

- wait time. Time spent waiting at the origin for a vehicle to become available.
- travel time
- delivery time reliability
- loss and damage
- packaginq cost
- handling cost
- tariff
- minimum shipment size requirements

A second important influence is the nature of the commodity being transported. The key variables which describe the commodity are:

- value. The price at the origin (the point of supply).
- shelf life. This is determined by spoilage or obsolescence.
- seasonality. Commodity demand may be seasonal or nonseasonal.

- density. Is the maximum shipment size determined by volume or by weight?
- perishability. This refers to the sensitivity of the commodity to environmental factors during transit.

A third influence is the state of the market for the commodity being ordered. This has special relevance to the choice of supplier (i.e., shipment origin). The key variables in this category are:

- price. The FOB factory price at each source of supply (including local wholesalers).
- quality. This is a difficult variable to measure, but it is often important.
- supply availability. Are orders filled from stock or from production runs?
- total volume of production of the supplier. Is this level of production compatible with the usage rate of the purchaser?

The fourth group of influential variables are the characteristics of the decision-maker's firm. This group includes the following:

- annual usage of the commodity being ordered
- variability in the usage rate
- consequence of a stockout. Do stockouts lead to a plant shutdown, a switch to a less efficient process, the loss of sales, or the postponement of sales?
- reorder cost

- storage cost. This includes the fixed cost of the warehouse and the variable cost of the crew.

- capital carrying cost

These key variables are summarized in Figure 2. It is clear that these variables interact in a complex manner. The next section addresses the problem of developing a theoretical framework that can be used to weave together the level of service attributes, commodity attributes, market attributes and receiver attributes.

Key Variables in Freight Demand

$$V_{ijmq}^k = f(T, C, M, R)$$

V = volume of freight flow  
 k = commodity type  
 i = origin  
 j = destination  
 q = shipment size  
 m = mode

T  
Transport Level  
 Of Service Attributes

wait time  
 travel time  
 delivery reliability  
 loss and damage  
 tariff  
 minimum shipment size  
 special services  
 packaging cost  
 handling cost

C  
Commodity  
 Attributes

value  
 shelf life  
 seasonality  
 density  
 perishability

M  
Market Attributes  
 (at potential origins)

price  
 quality  
 availability  
 production rate

R  
Receiver Attributes  
 (at the destination end)

use rate  
 variability in use rate  
 stockout situation  
 reorder cost  
 capital carrying cost  
 risk of stockout

## The Logic Behind Freight Transportation Decisions

How does a manager decide which mode, shipment size and supplier to choose? Classical microeconomics tells us that these decisions are made to maximize the profit of the firm, where profit is simply equal to revenue minus costs. However, a firm in a competitive market cannot influence the price of the good which it produces. Assuming that the size of the market is fixed and that the firm's market share is stable, then the firm's revenue will be fixed (in the short run). Under these conditions the objective of the manager is to minimize costs. Of course in the real world, revenues are not fixed. Yet many decision-makers do try to minimize costs subject to a profit constraint because this is a simple objective function to apply, whereas pure profit maximization is very complicated.

The total variable cost that is to be minimized is composed of three components: wages, purchase costs, and logistics costs. If the average daily rate of production is constant, then we may assume that wages are fixed. Thus the key costs affecting transportation decisions are the cost of purchasing supplies and the logistics costs. Given the volume of material needed, the cost of supplies will depend on the choice of a supplier and a shipment size. The annual purchasing cost for a single input process is simply the FOB factory price multiplied by the annual usage of the input.



The logistics cost has five components: ordering cost, transport cost, storage cost, capital carrying cost and stockout cost. These costs are functions of the level of service attributes, commodity attributes, market attributes and receiver attributes shown in Table 1, as well as the shipment decisions. The mathematical relationships are rather complex, and therefore the derivation of cost equations will be deferred to Chapter 6. However, the important interdependencies can be stated qualitatively in the following manner:

#### Ordering Cost

ordering cost per year = f(cost per order, frequency of orders)

frequency of orders = f(usage rate, shipment size)

This item represents the administrative cost of sending out orders. The cost of a single order is a receiver attribute. It is largely independent of the supplier, mode and shipment size decisions. However, the annual cost of ordering depends on the frequency of orders, which is a function of the shipment size decision and the receiver's annual usage of the commodity.

Transport Cost

transport cost per item = f(rate, other costs)

other costs = capital carrying cost, handling cost,  
packaging cost, loss, and damage, spoilage

capital carrying cost = f(wait time at the origin, travel time,  
commodity value, cost of capital)

spoilage = f(shelf life, wait time at the origin, travel time,  
commodity value)

This term includes all of the factors which are directly related to the cost of transporting the shipment. The transport cost depends primarily on the level of service attributes, which depend on the mode, shipment size and supplier decisions.

Storage Costs

storage cost per year = f(size of the safety stock,  
size of the non-safety stock,  
commodity density, perishability)

size of the safety stock = f(variability in use rate,  
reliability of delivery,  
chosen risk of stockout)

size of the non-safety stock = f(use rate, shipment size)

The annual cost of storing items at the receiver's plant depends on the size of the stockpile and the cost of maintaining each item. The storage cost per item is a function of the commodity attributes, such as shelf life, density and value. The size of the stockpile depends on the size of the safety stock and the non-safety stock. The size of the non-safety stock depends on the shipment size which is chosen. The size of the safety stock is determined by the risk of stockout chosen by the receiver and the variability in the inventory process. The variability in the inventory process is caused by the unreliability of the supplier, the unreliability of the carrier, and the variability in the receiver's daily usage.

#### Capital Carrying Cost for Items in Storage

carrying cost = f(size of the safety stock,  
size of the non-safety stock,  
commodity value, cost of capital)

The capital carrying cost reflects the opportunity cost of material tied up in the stockpile at the receiver's plant. This cost is a function of the size of the stockpile, the value of the commodity and the cost of capital (i.e. the interest rate). As explained above, the size of the stockpile is determined by the transport decisions, the chosen risk of stockout, and the variability in the receiver's daily usage.

Stockout Cost

stockout cost per year = f (stockout risk per order,  
frequency of orders,  
cost of a stockout)

frequency of orders = f (usage rate, shipment size)

cost of stockouts = f (value of the good being produced,  
consequence of stockouts)

The annual cost of stockouts depends on the cost of a stockout, the risk of stocking out during a reorder period, and the number of reorder periods which occur annually. The receiver can choose a stockout risk by adjusting the size of the safety stock. Given the risk of stocking out on any one order, the annual number of stockouts is determined by the frequency of orders (i.e. the shipment size decision). The stockout cost may be available for use in the model. If this is not the case, then the cost of a stockout can be estimated based on the value of the good being produced and other receiver attributes.

These five logistics costs and the purchase cost are the determining factors in the decision making process discussed in the previous section.

In choosing a supplier, a mode and a shipment size for an order, there are many tradeoffs available for the reduction of the purchase and logistics costs. Four of particular interest are:

- purchase price vs. transport cost
- large orders, low transport rate, high storage and carrying cost vs. small orders, more frequent stock-outs, high handling cost, low storage cost, high transport rate
- transit time vs. perishability
- reliability of delivery vs. high safety stock costs

It is important that a freight demand model capture these tradeoffs.

In the next chapter, sources of data are reviewed. Following that, the literature on freight demand models will be examined to determine the extent to which existing models include the key variables, functional relationships, and tradeoffs.

## Chapter 3

### A Review of the Available Data

In preparation for the literature review presented in Chapter 4, it is useful to compare the stock of available data with the list of key variables that was developed in the preceding chapter. Since almost all existing literature on freight demand modeling pertains to the rail, truck or barge modes, this review of data will focus on these three modes.

#### Sources of Intercity Flow Data

The most widely available source of information on intercity commodity flows is the Census of Transportation [3]. This census was first conducted in 1963 and it has been repeated with an expanded format in 1967 and 1972. It is important to note that the Census of Transportation is composed of three independent sections: the National Travel Survey, the Truck Inventory and Use Survey, and the Commodity Transportation Survey. The National Travel Survey is not relevant to studies of freight demand because it deals only with passenger transportation. The Truck Inventory and Use Survey contains information on fleet

---

3. Department of Commerce, Bureau of the Census, 1967 Census of Transportation, Vols. 1, 2, and 3, Washington, D.C., 1971.

characteristics and operations. These data are stratified by major use group (such as agriculture, mining, construction, manufacturing, etc.); however, no information is gathered on the use of particular types of trucks for hauls of particular commodities between origins and destinations. Because of the generality of these data, this survey is not directly useful in constructing freight demand models.

Specific data on commodity shipments is contained in the Commodity Transportation Survey. The purpose of this part of the Census of Transportation is to measure the intercity flow of all manufactured commodities. To facilitate this study, the universe of manufacturing activities has been divided into three broad segments. The first segment is the small industrial plant sector, which includes all plants with 10 to 19 employees. The survey of this segment consists of a random sample of 2000 plants drawn from the "universe" list of such plants. For each observation, the total value of products was used as a measure of the total volume of shipments. Data were collected on the annual use of each mode for the transport of each commodity produced. However data on individual shipments were not collected.

The second segment in the Commodity Transportation Survey consists of printing and publishing establishments, except newspapers and magazines. The universe of all such plants was divided into four plant size groups and a different sampling rate was used for each group. In total, about 1400 plants were

sampled. The same information was gathered for each observation as in the small plant sector described above.

The third segment in the CTS is the "Major Industrial Sector", which includes all manufacturing plants with 20 or more employees. This segment covers about 96 percent of all tonnage of intercity shipments of manufactured goods. For the 1967 Census, the surveying procedure in this segment began with a subdivision of plants into six volume of shipment categories based on the class of commodity produced by a plant and the average volume of production of plants in that class [4]. The 13,000 observations budgeted for the Major Industrial Sector were allocated among the six volume of shipment categories in proportion to the number of classes (called 'shipper classes' in the Census) in each volume category. Within each of the six volume of shipment categories, the selection of plants to sample was based on the number of employees in the plant and geographic location [5]. Thus, a list of about 13,000 plants was drawn up and from each plant a random sample of from 100 to 200 shipment records was chosen. This sample of about 1.4 million shipments formed the principle source of data for the 1967 Commodity Transportation Survey. A similar procedure was used to prepare

---

4. Census of Transportation data from previous years was used to assign shipper classes to one of the six volume of shipment categories.

5. For a more detailed explanation see "Sample Design, Commodity Transportation Survey, 1967 Census of Transportation", Bureau of the Census, 1968.



the 1972 Census.

The shipment records obtained from the three surveys described above have been extrapolated to the universe level. The universe level data is then published in three formats. The first format for the data consists of an aggregation into 85 shipper classes and 25 shipper groups. The second format consists of an aggregation into commodity groups by 3 digit STCC code. And the third format includes a presentation of the data by geographic areas: 25 Production Areas, 9 geographic divisions and selected states. The types of statistics published in the CTS are similar for all three formats. Data are shown for the tons and ton-miles of shipments by mode, length of haul, commodity group, shipment size, size of plant, origin or destination. To avoid disclosure of the activities at a single plant, the data is either cross-stratified by no more than three variables at one time, or it is withheld from publication.

Besides the three formats used for the publication of CTS data, the Census Bureau has prepared two public use computer tapes. These tapes contain data on origin to destination flows cross-stratified by commodity (2, 3, 4, and 5 digit STCC), mode (6 types), and shipment size (20 groups). One tape contains state to state flows and the other contains flows from 25 Production Areas to 55 Market Areas [6]. The Production Area to

---

6. The 1972 Census of Transportation tape is based on 27 Production Areas and 59 Market Areas.

Market Area tape represents the most disaggregate data available from the Census of Transportation. However these data are not disaggregated to the level of the individual decision-maker. Also many pieces of information have been withheld to avoid disclosure of shipments from individual plants.

Although the Census of Transportation contains a tremendous amount of information, it falls short of being an ideal data set for demand modeling applications for several reasons. First, the Census contains no information on the level of service associated with a shipment (i.e. rate, travel time, loss, damage, reliability of delivery, etc.). In theory, the level of service could be derived from tariff books, carrier's schedules, and other sources. However the Census data is aggregated enough that the level of service cannot be inferred unambiguously.

The second and most severe shortcoming of the Census of Transportation is that the origin-destination flow data do not include any information about the type of firms sending and receiving the shipments. As was argued in the previous chapter, the choices of supplier, mode and shipment size are integrally linked to the logistics cost functions of the firms involved. However this link cannot be studied with the Census data.

A third problem with the Census is the fact that it covers only part of the total freight transportation picture. Since only manufactured goods are included in the CTS, some very large tonnage flows of commodities such as coal, crude petroleum and

timber are not represented. The CTS also excludes flows of imports and transshipments. On top of all of this, many manufactured commodities are not represented in the most disaggregate Census statistics because of disclosure problems.

In spite of its shortcomings, the Census of Transportation is still the most detailed reference available on intercity shipping.

Another source of flow data is the Carload Waybill Statistics [7]. This report was published by the Interstate Commerce Commission until 1969, at which time responsibility was transferred to the Department of Transportation. The data base used to prepare the statistics is composed of a one percent sample of all rail waybills. This sample is collected simply by requiring railroads to turn in copies of all waybills with serial numbers which end with the digits '01'.

The Carload Waybill Statistics offer several features not found in the rail data published in the Census of Transportation. First, data is included on all types of commodities shipped by rail, not just manufactured goods. This information is available at the 2, 3, 4, and 5 digit STCC levels. Secondly, this source includes data on revenue per ton-mile (i.e. rate) which is an important level of service characteristic. Also included are data on shipment size and length of haul, as well as the total

---

7. Department of Transportation, Carload Waybill Statistics, 1973, Washington, D.C., 1974.

volume of flow in tons and ton-miles.

For purposes of estimating demand models, the main shortcoming of the Carload Waybill Statistics is the lack of geographic detail. Origin to destination flows are reported by commodity in two forms: state to state flows, and flows between the five territories defined by the I.C.C. [8]. Unfortunately, no city to city flow data is published. Hence the data on shipment size, revenue, and length of haul are of marginal use because they are averaged over such a wide range of O-D pairs. Nevertheless, the Carload Waybill Statistics are important because they are the most detailed source of data on rail shipments of raw materials.

A third source of shipment data is Waterborne Commerce of the United States, which is published by the Army Corps of Engineers [9]. The data used to prepare these statistics are collected on special reporting forms which each carrier is required to submit. The published information covers inland and coast-wise movements of domestic shipments of all commodities.

Parts 1 through 4 of Waterborne Commerce cover activities in ports and on waterways in various sections of the country. Part 5 contains the national summaries. Of particular interest is

---

8. The five territories are: Official, Southern, Southwest, Mountain Pacific and Western Trunk Lines.

9. Department of the Army, Corps of Engineers, 1973 Waterborne Commerce of the United States, Parts 1,2,3,4, and 5, Washington, D.C., 1974.

Section 2 of Part 5 which contains origin to destination flow data. The origins and destinations are defined in terms of rivers (segments), canals and major ports. For each O-D pair, the volume of shipments in tons is presented for selected raw materials and manufactured goods at the 4 digit STCC level of detail. Unfortunately, no data on shipment size, revenue or travel time are presented in this section.

There are several other sources of flow data which may be of use in some freight demand studies. However, these reports suffer from the drawback that they contain no information on origin to destination flows. The Interstate Commerce Commission publishes the "Freight Commodity Statistics - Motor Carriers of Property" and "Freight Commodity Statistics - Class I Railroads" [10][11]. These reports contain data on the volume of flow and the revenue associated with shipments of regulated commodities (described by 2,3,4 and 5 digit STCC). The rail data is available for each of three districts, and also at the national level. The truck data is compiled for nine regions, and for the nation as a whole.

---

10. Interstate Commerce Commission, Bureau of Accounts, "Freight Commodity Statistics - Motor Carriers of Property", Washington, D.C., published annually.

11. Interstate Commerce Commission, Bureau of Accounts, "Freight Commodity Statistics - Class I Railroads", Washington, D.C., published annually.

The Interstate Commerce Commission also publishes a set of reports titled "Transport Statistics in the United States" [12]. These reports cover all of the regulated carriers. They are focused primarily on the finances of the carriers and the stock of transportation equipment. However, the barge report contains national level data on flows and revenues for regulated commodities described by 2,3,4, and 5 digit STCC.

The carrier's associations also publish summaries of freight flows. The American Trucking Association report titled "American Trucking Trends" contains information on the volume and revenue of truck shipments [13]. Most of this data pertains to a combination of commodities at the national level. Nevertheless, some data is listed for commodity groups and regions of the country. The Association of American Railroads publishes a similar report titled "Yearbook of Railroad Facts" [14]. This report contains some information on individual railroad companies, but very little data on individual commodities. In summary, the I.C.C., A.T.A. and A.A.R. publications are useful only for studying general freight transportation trends at the national level.

- 
12. Interstate Commerce Commission, Bureau of Accounts, "Transport Statistics in the United States", published in six parts: Part 1 - Railroads; Part 5 - Carriers by Water; Part 6 - Pipelines; Part 7 - Motor Carriers; Part 8 - Freight Forwarders; Part 9 - Private Car Lines, Washington, D.C., published annually.
13. American Trucking Association, "Transportation Facts and Trends - 1974", Washington, D.C., published annually.
14. Association of American Railroads, "Yearbook of Railroad Facts - 1974", Washington, D.C., published annually.

In assembling a picture of the intercity freight transportation system there is one notable piece of information which is missing. At this time, there is no source of data on origin to destination flows of raw materials and agricultural commodities carried by truck. A small amount of information is available in a Department of Agriculture publication titled "Fresh Fruit and Vegetable Unload Totals" [15] this source contains data on the flow of agricultural commodities into 41 major cities. However it does not include any data on shipment origins. Another possible source of data is the Truck Commodity Flow Study conducted by the Federal Highway Administration in 1973. But the results of this study have not been published yet.

#### Sources of Market and Receiver Attributes

The primary source of information concerning markets and firms are the Economic Censuses: the Census of Manufacturers, the Census of Wholesale Trade, the Census of Retail Trade, the Census of Selected Service Industries, the Census of Agriculture and the Census of the Mineral Industries. These publications contain production related statistics for individual commodities at the 2, 3, and 4 digit SIC levels of detail. The list of pertinent data contained in these publications includes the number of

---

15. Department of Agriculture, Consumer and Marketing Service, "Fresh Fruit and Vegetable Unload Totals", Washington, D.C., 1974.

plants producing a commodity, the value and volume of output, and the number of employees. This data is available for both states and large SMSA's.

Despite the fact that they contain large quantities of data, the Economic Censuses are not well suited for use in freight demand modeling. One problem is that they are designed to measure outputs, while it was argued in Chapter 2 that the demand for freight transportation is determined in large part by decisions concerning the ordering of inputs. Also, there is no way that the Economic Censuses can be used to link the production activities of a firm of a particular size with the transportation used by that type of firm for the shipment of either inputs or outputs. This makes it impossible to test the simple hypothesis that large, high volume firms favor large shipment sizes, while small, low volume firms favor small shipment sizes. The severity of this problem will be discussed in more detail in Chapter 7.

#### Sources of Level of Service Attributes

There are no publicly available data sets which give extensive information on the level of service for freight shipments. Many researchers in freight demand modeling have found it necessary to estimate level of service data using supply models. However, the selection of available supply models is also disappointing.



In theory, tariff books could be used to look up rates for the shipments described in sources such as the Census of Transportation. In practice, this is very difficult because tariffs are voluminous and constantly changing. Furthermore, one has to be a rate expert to use the tariffs accurately. For these reasons, most researchers have used some form of rate estimation equation. Some rate models have been estimated using the average revenue and average length of haul data presented in the Carload Waybill Statistics. Other models have been based on rate data contained in a paper by Alexander Morton [16]. The results of these efforts have been poor because of the aggregate nature of the data used for calibration. New work in this area will be discussed in Chapter 7.

As in the case of rates, the lack of travel time data has forced many researchers to develop supply models. Most of the travel time models discussed in the freight demand modeling literature have involved estimation of the mean travel time as a function of the length of haul and the number of intermediate terminals or yards. However it was shown in Chapter 2 that logistics decisions are a function of the entire distribution of travel times, and not just the mean. One of the more interesting models for estimating travel time distributions has been

---

16. Morton, A.L., "Competition in the Intercity Freight Market", Office of Systems Analysis, Dept. of Transportation, 1971.

developed as part of the Railroad Reliability Project at M.I.T. [17]. However this model has not been used extensively to date.

Loss and damage data is so scarce that there have been very few attempts to develop supply models for these level of service attributes [18]. The main source of data on the loss and damage of truck shipments is the quarterly report published by the I.C.C. [19]. The data in this report is stratified by region, cause and commodity (at the 3 digit STCC level). Another source of data on truck shipments is Trinc's Blue Book of the Trucking Industry, which gives the annual cost of L/D incurred by each carrier.

The best source of information on the loss and damage of rail shipments is the annual L/D report published by the Association of American Railroads [20]. These data are classified by cause and commodity (2 digit STCC level) for both carload and LCL shipments. In general, this rail data and the

-----  
 17. The Industry Task Force on Reliability Studies, and the Center for Transportation Studies at M.I.T., "Railroad Reliability and Freight Car Utilization: An Introduction", Massachusetts Institute of Technology, 1975.

18. One model of loss and damage has been proposed by Allen (1973). However this model has never been implemented. Another model is presently under development at the Center for Transportation Studies at M.I.T. .

19. Interstate Commerce Commission, "Freight Loss and Damage Claims, Motor Carriers of Property", published quarterly.

20. Association of American Railroads, Operations and Maintenance Department, Freight Claims Division, "Freight Loss and Damage", published annually.

truck data are too aggregate to reflect the loss and damage for shipments of particular commodities between particular origins and destinations.

In summary, most of the available information on freight transportation is oriented toward giving a broad picture of the situation at the national level. Very little data is available on the characteristics of individual shipments and shippers. As will be shown in the following chapter, the state of the data is largely responsible for the concentration of demand modeling research in the area of aggregate models.

## Chapter 4

## Literature Review

The purpose of this chapter is to review previous research in freight demand modeling. As mentioned in Chapter 1, relatively little work has been done in the area of freight demand modeling in comparison to the extensive body of literature on modeling 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. The freight demand models developed in these studies generally address one of three types of commodity flows: intra-city, inter-city, or international. Although the modeling methodology developed in this thesis is applicable to all three types of flows, the literature review is focused on models of inter-city shipping. Furthermore, the scope of this literature review has been narrowed to include only those studies dealing directly with models. There are a large number of reports dealing with the general structure and functioning of the freight transportation system which are not mentioned in the

---

21. See for example: "Competition Between Rail and Truck in Intercity Freight Transportation", Charles River Associates, December, 1969.

following discussion [21]. A few other reports containing projections of commodity flows have also been excluded because the work was based mostly on expert opinion rather than mathematical models [22].

In reviewing freight demand models, primary consideration has been 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 used in reviewing models 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

---

22. See for example: "Forecasts of Traffic and Revenues 1974-1980", prepared for the USRA by Temple, Barker and Sloane Inc., October 1974.

applicability of a model to a wide range of demand related freight transportation problems.

Demand models can be separated into two general groups: aggregate and disaggregate. These two groups of models are substantially different and therefore the freight demand models in these groups have been reviewed separately. Furthermore, in the following, the aggregate and disaggregate models have been grouped according to their dependent variable to facilitate comparisons between models designed for roughly the same kind of forecasting.

Table 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, 68	rate		production indices	log linear	7,8
	Tihansky, 64			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, circuitry 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
	Kresge, 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, 36	rate, travel time reliability, distance	value	annual volume	logit	1, 2, 13

Table 1 (continued)

Type of Model	Modeler	Level of Service Variables	Commodity Descriptors	Other Variables	Functional Form	Principal Data Sources
Disaggregate Mode Choice	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
	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 numbers after the modeler's name refer to the bibliography on page 192.

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



## 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.

### Aggregate Models of Intercity Freight Demand

Most of the freight demand studies done to date have utilized aggregate data from government sources. As described in Chapter 3, aggregate data on national, inter-regional, and interstate flows by commodity and mode are readily available. 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 first problem with models based on aggregate data is that the estimates of the model parameters are dependent on the system of aggregation used to prepare the data. This means that aggregate models are not likely to be transferable. In other words, aggregate models do not perform well when they are used to forecast the demand for regions, markets, commodities, and/or modes other than those included in the data used to estimate the model. Moreover an aggregate model may be completely invalidated by changes in the structure of the economy and transportation system over time. This is a severe limitation on the flexibility of this type of model.

A second problem with aggregate models is that they often lack many policy sensitive variables. Basically, the process of aggregation destroys much of the variability (i.e., explanatory power) of the data. For example, the average unreliability of transit time between two cities may be quite small, while the variation in arrival times observed by a particular receiver may

be quite large. Under these circumstances, the aggregate measure of unreliability would fail to explain why receivers are dissatisfied with the level of service. And because an aggregate measure of unreliability is often a poor explanatory variable, it is likely to be removed from the model. Similarly, many other policy sensitive level of service variables may appear to be insignificant in an aggregate model.

The third problem with aggregate models is a direct result of the shortcoming discussed in the previous paragraph. Since aggregate data often fail to describe the shipment alternatives as they are seen by the decision-maker, aggregate models usually fail to explain the decision making process in an intuitively clear manner. In other words, the mechanism that controls the demand for various forms of freight transportation is often difficult to deduce from a casual inspection of the model. The coefficients usually reflect the influence of many different factors which are not explicitly represented in the model. For this reason, aggregate models are of limited use as explanatory tools.

Due to these problems, 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 more complex models. To avoid confusion on this point the single choice and multi-choice models are reviewed separately in the following discussion.

### Aggregate Mode Choice Models

One of the best known studies of freight demand was conducted by Eugene Perle (1964). Perle postulated 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 modal 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 began by estimating a national level model using dummy variables for the commodity group, region and year. His model is of the following form:

$$\log(V_{m1} / V_{m2}) = \beta_0 + \beta_1 \log(r_{m1} / r_{m2}) + \sum_{i=1}^9 c_i R_i + \sum_{j=1}^5 d_j Y_j + \sum_{k=1}^5 f_k C_k$$

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)

$Y_j$  = (1 for year  $j$ , 0 otherwise)

$C_k$  = (1 for commodity  $k$ , 0 otherwise)

Perle estimated this model using ordinary least squares regression. The results were poor. 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 this 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 discussed in Chapter 2. 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 it 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 Edward 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 availability 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:

$$\left( \frac{V_{m1}}{V_m} \right) = \beta_0 + \beta_1 (r_{m1}) + \beta_2 (\text{rail availability})$$

where  $V_{m1}$  = volume carried by rail

$r_{m1}$  = 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, reliability, 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 Vasant Surti and Ali 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:

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

where  $V_{m1}^k$  = volume of commodity k carried by truck

$V_{m2}^k$  = 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 Surti/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:

$V_{ij}^k$  = 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

$c$  = circuitry 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:

$$\sin^{-1} \sqrt{\frac{V_{ij,m1}^k}{V_{ij,m1}^k + V_{ij,m2}^k}}$$

$$= \beta_0 + \beta_1 \log(V_{ij}^k) + \beta_2 \log(v) + \beta_3 \log(d) +$$

$$+ \beta_4 \log(L) + \beta_5 \log(c) + \beta_6 (B) + \beta_7 (S)$$

where  $V_{ij,m1}^k$  = volume of  $k$  carried from  $i$  to  $j$  by barge

and  $V_{ij,m2}^k$  = 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) [23].

-----

23. The Bureau of Economic Analysis has divided the U.S. into 173 zones. The Census of Transportation, the Carload Waybill Statistics, and the Waterborne Commerce data were retabulated for use in this zone system.

The results from estimating this model were mixed. The  $r$  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. Under 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:

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

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

$$AVC_m = \text{rate}_m + \beta_2 (\text{time}_m * \text{value}_m) + (\beta_3 / [V_{ij}^k]^{0.5})$$

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. Thus, this model can be considered to be a (partial) implementation of the conceptual model of decision making which was introduced in Chapter 2.

It should also 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. All of these features set this model apart as a distinctly different kind of policy analysis tool than those models previously discussed.

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  $r^2$  statistics were above 0.80. These encouraging results tend to support the opinion that this Mathematica model was a step in the right direction. Several features of this model are incorporated in the disaggregate demand model which will be proposed in Chapter 6.



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 utility of two competing modes. The form of this model is the following:

$$\frac{V_{ij,m1}^k}{V_{ij,m1}^k + V_{ij,m2}^k} = 1 / (1 + w)$$

where:

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

$$u = \left[ b \cdot \ln(v) + b \cdot v \right] \left[ b \cdot \ln(V_{ij}^k) + b \cdot (V_{ij}^k) \right]$$

$t_m$  = 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

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

This model was estimated using the same data base as was described in the preceding review and it 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.

Brian 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 highway 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:

$$\log \left( \frac{V_{m1}^k}{V_{m2}^k} \right) = \beta_0 + \sum_i \beta_i x_i$$

where  $x_i$  is an explanatory variable

and  $V_{m1}^k$  = volume of commodity  $k$  carried by rail

$V_{m2}^k$  = 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 the other independent variables for the  $x$ 's in the formula shown above.

This makes it somewhat more difficult to interpret the coefficients. On the other hand, Kullman included in his model a commodity attribute (value) and a level of service attribute (travel time reliability) which were not included in the first few models reviewed in this chapter.

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 rail rates used in Kullman's study were estimated using data from the 1967 Carload Waybill Statistics. The truck rates were estimated from data presented in a paper by Alexander Morton (1971). Rail transit time distributions were estimated from data made available to Kullman by the Penn Central Railroad. Mean travel times for truck were calculated as a function of the highway distance. Commodity values were estimated from data published in the 1967 Census of Manufacturers.

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. The decision hierarchy introduced in Chapter 2. Gives some support to the concept of a sequential model system. However, the choice hierarchy also includes feedback from short run decisions to long-run decisions which has not been adequately modeled in the systems which have been developed to date. Furthermore, the choice hierarchy includes joint decision making at some levels. None of the systems reviewed below has taken this into account.

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 [24]. 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 in this section, although they are referred to in the summary table at the beginning of the chapter.

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

---

24. A.D. Little, Domestic Waterborne Shipping Market Analysis, a study for the Federal Maritime Administration, 1974.

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 inter-regional 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 must be used to make the totals consistent. Mathematica developed an adjustment model using Lagrange multipliers. The objective of the Lagrangian is to minimize the flow 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 performance 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 [25]. The independent variables in this model are the changes in zonal population and employment.

Adjustments in modal split are made using a share model of

-----  
25. For a more complete explanation of the Fratar model see: Martin, B.; Memmott, F. and Bone, A., Principles and Techniques for Forecasting Future Demand for Urban Area Transportation Travel, M.I.T. Press, 1963.



the following form:

$$V_{ij,m}^k / V_{ijn}^k = (\beta_{1,m} t_{m1}^{\beta_{2,m}} r_{m1}^{\beta_{3,m}}) / (\beta_{1,m} t_{m1}^{\beta_{2,m}} r_{m1}^{\beta_{3,m}})$$

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 modal 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

As discussed in the previous section, 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 choice hierarchy discussed in Chapter 2 indicates that 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. In terms of their use in policy analysis, the aggregate joint demand models developed to date have shed little light on the decision-making process.



A joint aggregate demand model was estimated as part of Perle's (1964) study which was described earlier in this chapter. 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:

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

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)

$Y_j$  = (1 for year j, 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 James 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, inter-provincial 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 intra-regional 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 I 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 were 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 attributes 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 Lloyd Antle and Richard Haynes (1971) at the Institute for Water Resources. The data set used in this study is described in detail in Appendix I. 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  $Z$  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 James Hartwig and William Linton (1974). The data which they used is described in Appendix I. 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. An approach to this next step is discussed in Chapter 6.

## Chapter 5

### A Brief Review of the Theory of Disaggregate Demand Models

In the preceding chapter numerous references were made to the advantages of disaggregate demand models. Although several examples were briefly discussed, the exact nature of disaggregate models may not be clear to all readers at this point. However, a working knowledge of this type of model is assumed in the following chapters. In particular, the disaggregate approach is an integral part of the freight demand model which is proposed in Chapter 6. Therefore a short digression will be made in order to describe the fundamentals of disaggregate models. A more in-depth discussion can be found in a number of references, including Ben-Akiva (1973) and Charles River Associates (1972).

#### Attributes of Disaggregate Demand Models

A freight demand model is most likely to be used to analyze the impacts of policies on various segments of the freight market. The focus of this type of analysis is on the demand generated by groups of shippers and not the behavior of individual firms [26]. However if data describing each shipper

is available, then either of two modeling approaches may be employed. One approach is to combine the data on shippers in each group, and then estimate a model based on the group averages. Forecasts for groups of shippers can then be prepared directly from the estimated model. The alternative approach is to estimate a model based on the behavior of individual shippers. The estimated model can be used to forecast the demand generated by each shipper, and then these forecasts can be summed up to predict the demand associated with each group. This second approach is known as disaggregate modeling.

Although disaggregate modeling may appear to be indirect and clumsy, it has several very important advantages. One key feature of disaggregate models is their efficient use of data. If a model is estimated with one thousand aggregate data points, then one thousand sets of group averages must be prepared. If there are twenty-five shippers in each group, then the total data base must cover twenty-five thousand shippers. On the other hand, if a disaggregate model is estimated with one thousand data points, then the data base need only cover one thousand shippers. Furthermore, the disaggregate model estimated on a data base of one thousand observations will predict demand at least as accurately (and usually more accurately) than the aggregate model

-----  
26. In this chapter the term 'shipper' will be used loosely to refer to any party involved in making shipment decisions. Note that in all other chapters the shipper is the party located at the shipment origin.

estimated on a data base of twenty-five thousand shippers. The reason for this is that the aggregation of data before model estimation results in the loss of much of the information in the data. In most cases the average characteristics of firms in one group will be similar to the average characteristics of firms in other groups, despite the fact that the firms within any one group may differ significantly. Thus, aggregated data usually fails to reflect the variability in shipper attributes which is so important in explaining the variability in shipper behavior.

Another feature of disaggregate models is that they are transferable. In an aggregate data set, each data point represents the mean of the distribution of firms in a group. If a model is estimated with aggregate data, then the values of the model parameters depend implicitly on the distributions of firms in each group. Hence the model cannot be used for forecasting in another situation where the distributions within the groups are not the same. However, since disaggregate models are estimated with data on individual shippers, this kind of model can be applied in areas other than those represented in the estimation data set. In addition, disaggregate models can be used to forecast the demand generated by groups of any size. Thus a single disaggregate model might supplant aggregate models at several planning levels.



### The Structure of Disaggregate Models

At the level of the individual plant manager, freight shipment decisions can be characterized as a selection from a set of alternatives. This set of alternatives is often referred to as a 'choice set' and it is composed of a list of all of the available options. For example, in the simple case of a mode split model the choice set might be defined as follows:

alternative 1-Rail

alternative 2-Truck

alternative 3-Barge

alternative 4-Air

We can denote this set with the symbol  $A_t$  where  $t$  indicates that we are referring to the set of alternatives available to the  $t^{\text{th}}$  manager in the study.

Associated with each alternative is a generalized cost. The cost to manager  $t$  of choosing alternative  $i$  may be denoted by  $C_{it}$ . Given that one and only one alternative is selected from the choice set, then alternative  $i$  will be selected if and only if :

$$C_{it} \leq C_{jt} \quad \text{for all } j \text{ in } A_t \quad (1)$$

In other words, shipment method  $i$  will be chosen if and only if its cost is less than or equal to the cost of all alternatives which are available.

The generalized cost of an alternative is a function of the attributes of that alternative. As discussed in Chapter 2, there are two types of attributes which directly affect the cost of an alternative. These two types are transport level of service attributes (e.g. transit time, tariff, loss and damage, etc.) and market attributes (especially prices of goods in different areas). However the cost of any given alternative will be evaluated differently by decision-makers involved in different kinds of activities. Therefore the generalized cost of a freight transport alternative should be expressed in the model as a function of the commodity attributes and the characteristics of the decision-maker as well as the level of service attributes and market attributes. The exact specification of the cost function will be addressed in the next chapter.

Due to measurement errors, unobservable information, and other deficiencies in the available data, it is usually impossible to calculate the exact cost of each alternative. Thus the cost can be expressed as the sum of two components, as follows:

$$C_{it} = c_{it} + \epsilon_{it} \quad (2)$$

where  $c_{it}$  is the observable part of the cost function and  $\epsilon_{it}$  is

an unobservable random element. If the generalized cost associated with each alternative includes a random component, then the deterministic choice criteria given in equation (1) based only on the observed cost component,  $c_{it}$  will not always predict the actual choice. Under these circumstances, only the probability of choosing each alternative can be predicted, rather than a deterministic prediction.

The probability that alternative  $i$  will be selected by manager  $t$  equals the probability that the cost of alternative  $i$ ,  $C_{it}$  is smaller than or equal to the cost of all other available alternatives. This can be expressed formally in the following manner:

$$P(i:A_t) = \text{Prob} [ C_{it} \leq C_{jt}, \text{ for all } j \text{ in } A_t ] \quad (3)$$

where  $P(i:A_t)$  is the probability of manager  $t$  selecting alternative  $i$  from his choice set  $A_t$ . Substituting equation (2) into equation (3) produces the following expression:

$$P(i:A_t) = \text{Prob} [ (\epsilon_{it} - \epsilon_{jt}) \leq (c_{jt} - c_{it}), \text{ for all } j \text{ in } A_t ] \quad (4)$$

This expression implies that the joint probability distribution of the random components determines the form of the model which relates the systematic cost functions to the choice probabilities.

One specific assumption about the joint distribution of the

random elements leads to the multinomial logit model, which is the only probabilistic choice model that has been extensively applied to problems involving more than two alternatives. The random elements are assumed to be independently and identically distributed according to the Weibull distribution:

$$P(\epsilon \leq w) = e^{-ne^{-w}}$$

where  $n$  is any positive constant. Substituting this distribution into equation (4) and integrating will result in the logit model which can be written in the following manner:

$$P(i:A_t) = \frac{e^{c_{it}}}{\sum_{j \in A_t} e^{c_{jt}}}$$

The systematic cost functions are usually restricted to be linear in the parameters:

$$c_{it} = \sum_{k=1}^K (x_{itk} * \beta_k)$$

where each  $x_{itk}$  is an independent variable describing alternative  $i$  and decision-maker  $t$ .

The independent variables in the model can be formulated as either alternative specific or generic. An alternative specific variable is associated with a different coefficient in the cost

function of each alternative. Thus it is possible to have a variable such as travel time unreliability affect the cost of the rail alternative more than the barge alternative. It is also possible to constrain the coefficient of travel time unreliability to be zero in the barge cost function, if this variable has no effect on the cost of using barge.

Generic variables have the same coefficient in all cost functions, although the value of the variable may vary from alternative to alternative. In fact, a generic variable cannot have the same value for all alternatives. For this reason, the presentation of a model specification in the next chapter includes comments on how the value of each generic variable varies across the set of alternatives.

It should be noted that generic variables have the distinct advantage of being 'alternative abstract'. This means that it is possible to use their coefficients to construct a cost function for a new alternative without re-estimating the model. Obviously this is an important feature if the demand model is to be used to evaluate policies involving new technical innovations and new services.

#### Data Considerations and Model Estimation

To estimate a disaggregate freight demand model, the data base must contain the attributes of all of the alternatives, as well as the commodity attributes and the variables describing the

decision-maker. Note that the data must describe both the chosen and unchosen alternatives. The observed dependent variable is assigned the value of one for the chosen alternative and zero for all others. However the forecasts produced by the model are probabilities which will range between zero and one for each alternative in such a way that the following condition will always hold:

$$\sum_{i \in A} P(i:A)_t = 1$$

The estimation technique which is often used is the maximum likelihood method. When the logit model is estimated using this method there are no limitations on the number of variables or on the number of alternatives. Furthermore, the number of alternatives need not be identical for all observations.

#### The Need for Further Research into the Formulation of Disaggregate Models

One of the major advances in demand modeling in the last five years has been the development of joint choice disaggregate models. These models have the same properties and functional form as the disaggregate models described in the preceding sections. However, each alternative in the choice set of a joint model represents a combination of alternatives for a set of two or more choices. The choice hierarchy discussed in Chapter 1 indicates that in freight transportation the choices of mode and

shipment size are often made jointly. The choice of a supplier may also be included in this set. Hence, disaggregate joint choice models would appear to be of particular importance in modeling freight demand. However, none of the existing models is well suited to this task. The problem arises from the fact that the set of mode and supplier alternatives is discrete, while the shipment size alternatives are continuous and infinite.

The multinomial logit model is the only disaggregate model which has been used extensively to model joint choices. This model requires that the choice set contain a finite number of alternatives. The only way to address the choice of shipment size is to divide the range of sizes into a set of discrete segments. If the segments are made too small, then the resulting model will violate the assumption of independence between the random elements in the cost functions of different alternatives. This can lead to serious biases in the estimated coefficients. On the other hand, if the segments are made too large then it becomes difficult to describe the alternatives, especially with variables such as the transport rate and the FOB price. Unfortunately it is very hard to find any solid middle ground between these pitfalls. It appears that a new type of model is needed to solve this problem.

A statistical model capable of handling the joint choice of a discrete and a continuous variable was described in a recent paper by Richard Westin (1975). The modeling approach developed

by Westin consists of five steps. In terms of the joint choice of a mode and a shipment size, these steps are the following:

1. Define a multivariate probability distribution of the desired shipment sizes associated with all of the available modes.

2. Define a model of the probability of choosing each mode given the desired shipment sizes on all modes.

3. Combine the results of steps 1 and 2 to obtain the joint probability distribution for all modes and shipment sizes.

4. For each decision-maker, integrate the joint probability distribution over all unchosen shipment sizes to obtain the marginal probability distribution of the chosen mode and shipment size.

5. Combine the results of step 4 for all individuals in the sample to obtain the likelihood function from which the model parameters can be estimated.

Westin has demonstrated this technique using a probit model for the discrete choice variable and a bivariate normal distribution for the continuous choice variable [27]. It should be noted that this approach is still in the developmental stage and has never

-----

27. Westin has also shown that a logit model may be used in place of the probit, although this complicates the mathematics considerably.



been empirically tested.

The model proposed by Westin is interesting, but it suffers from two major shortcomings. The first problem is that the discrete choice has been limited to two alternatives. Thus it is not (currently) possible to address the choices of a mode, a supplier, and a shipment size in the same model. Furthermore, this restriction makes it difficult to apply the model in the vast majority of freight transportation cases which involve several modal options. On the other hand, there is no theoretical reason why Westin's approach could not be extended to overcome this problem. The key question is whether the extended model can be made mathematically tractable.

The second and more severe problem with the current version of Westin's model is the very restrictive way in which the continuous choice variables are allowed to enter into the cost functions associated with the discrete choice variables. In the case of a mode and shipment size model, the desired shipment sizes can only enter linearly into the cost function of any of the modal alternatives. Thus, it would be impossible to use the transport rate to explain the choice of mode because the rate is a non-linear function of the shipment size. Since the shipment size affects many logistics cost factors in a non-linear manner, Westin's model is (currently) of limited usefulness in freight demand modeling. Nevertheless, Westin's work is an important contribution. Continued research may lead to methods of

generalizing the model to make it more useful.

It is evident that none of the available disaggregate demand models fill all of the requirements for a joint choice model of the demand for freight transportation. In the next chapter a specification for a joint choice model is presented. This specification is general enough to be applicable to a wide range of disaggregate models. Of course it will have to be refined somewhat when a suitable model has been developed. In the meantime, the specification should be helpful in guiding research into new functional forms for disaggregate models.

## Chapter 6

### Development of a Model Specification

As was pointed out in Chapter 4, the freight demand models which have been developed to date have three major shortcomings. First, they do not take into account the fact that some decisions such as those concerning mode and shipment size are made jointly. Secondly, these models lack many of the level of service attributes, commodity attributes, receiver attributes and market attributes which were discussed in Chapter 2. And thirdly, most models have failed to capture the relationships between these key variables. In this chapter a model specification is presented which utilizes the theory of logistics management to overcome these three pitfalls. This specification is designed for a disaggregate choice model of the type described in the previous chapter.

#### The Scope of the Model

In specifying a model, the first issue that must be addressed involves the scope of the model. Which shipment decisions will be predicted by the model? And which decisions will be part of the given conditions? To answer these questions it is helpful to review the discussion of the decision hierarchy

which was included in Chapter 2. It was argued in that chapter that the short run choices of mode and shipment size are almost inseparable. Shipment size strongly influences the level of service provided by each mode, and conversely the choice of mode restricts the set of feasible shipment sizes. After mode and shipment size, one of the most flexible decisions involves the choice of a supplier (i.e. shipment origin). This decision affects not only the relative quality of service provided by each mode, but also the availability of many modal alternatives, especially barge and air transport. Thus the interdependencies between the mode, shipment size and supplier decisions make it desirable to model these three choices together.

In theory it is desirable to build a model which encompasses the long range decisions on the plant location and size, as well as the three short(er) run decisions discussed above. From a practical point of view, this may not be possible because long run decisions are influenced by a wide range of non-transportation variables, above and beyond the transportation related variables. The plant size decision is essentially the choice of a long run average production rate. This decision is influenced by the supply-demand market equilibrium of many commodities. Similarly, the location decision is influenced by many characteristics of markets in various regions of the country. However it is unlikely that the choices of mode, shipment size and supplier would affect the plant size or

location in the short run because of the large capital investment involved. For short run policy analysis, it is reasonable to model the joint choice of mode, shipment size, and supplier, while viewing the plant size and location as fixed.

It should be noted that the assumption of a fixed plant size (i.e. a constant long run average production rate) does not imply that plant production is the same every day. Daily fluctuations in plant production and the resulting fluctuations in the usage rate of inputs are an important influence on short run transportation decisions. In the specification developed in this chapter, these fluctuations are represented by a probability distribution of the daily usage of each input. Based on the assumption of a constant long run average production rate, it has been assumed that the distribution of daily usage rates for each buyer located at  $j$  with usage rate distribution  $g$  and ordering input is invariant over the short run.

commodity  $k$  from a supplier at  $i$  in quantity  $q$  to be transported by mode  $m$  is a function of transport level of service attributes, commodity attributes, market attributes, and attributes of the receiver or buyer. Thus the set of alternatives associated with a choice model of this type is of the following form:

alternative 1-Buy in city 1 in quantity 20 and ship by truck-TFL  
 alternative 2-Buy in city 3 in quantity 90 and ship by rail-FCL  
 alternative 3-Buy in city 7 in quantity 45 and ship by truck-FTL

alternative  $n$ -Buy in city  $i$  in quantity  $q$  and ship by mode  $m$

Before proceeding with a detailed development of the independent variables, it may be helpful to summarize the type of model which is being proposed in this chapter. The model may be written symbolically in the following manner:

$$\text{prob}(i,m,q, \text{ given } j,u,k) = F(T,C,M,R)$$

where k = commodity being ordered  
 i = shipment origin  
 m = mode  
 q = shipment size  
 j = shipment destination  
 u = use rate distribution for commodity k  
 T = transport level of service attributes  
 C = commodity attributes  
 M = market attributes  
 R = receiver attributes  
 F = functional form for the choice model

The above expression simply states that the probability of a buyer located at j with use rate distribution u ordering commodity k from a supplier at i in quantity q to be transported by mode m is a function of transport level of service attributes, commodity attributes, market attributes, and attributes of the receiver or buyer. Thus the set of alternatives associated with a choice model of this type is of the following form:

alternative 1-Buy in city 1 in quantity 20 and ship by truck-LTL  
 alternative 2-Buy in city 3 in quantity 90 and ship by rail-FCL  
 alternative 3-Buy in city 7 in quantity 45 and ship by truck-FTL

·  
·  
·  
·

alternative n-Buy in city i in quantity q and ship by mode m

Given this set of alternatives, the plant manager's decisions are motivated by a desire to minimize the sum of the purchase cost and the logistics cost. As explained in Chapter 5, choice models such as the logit model are based on the principle of cost minimization. Thus, the central issue in the specification of the model is the development of a mathematical representation of the key costs.

#### Specification of the Independent Variables

In Chapter 2, logistics management theory was used to develop a list of the cost factors which influence the mode, shipment size and supplier decisions. Equations for these cost factors are available in the logistics management literature. These equations could be included directly in a disaggregate freight demand model. But this would require excessively detailed data for estimating the model and forecasting. In the following, an attempt is made to construct variables which capture the same effects as the logistics cost equations, but without the use of as much detailed data.

One other point should be kept in mind when specifying cost variables for a demand model. All costs should be calculated on a per unit of input, or per year basis. Costs calculated on a per shipment basis should not be used. The reason for this is simply that a large order is always more expensive than a small

order, although the large order may be more economical. In this chapter all costs are calculated on a per unit basis. The annual costs can be calculated in a similar manner.

The following variables represent the principle logistics costs associated with the choices of supplier, shipment size and mode:

#### Purchase cost per unit

$B$  (FOB price at origin  $i$ )

1

This variable is defined as the FOB price of the commodity being purchased. In most cases this price will vary from supplier to supplier and it will also depend on the size of the order. However, the price will usually be unaffected by the choice of mode.

The coefficient  $B$  serves as a scaling factor. Thus it can be used to normalize all of the other coefficients. If this is done, then the resulting coefficients will be expressed in units which are compatible with costs measured in dollars.



Ordering and handling costs per unit

$$\frac{B \text{ (orders/year)}}{2} / (\text{annual usage})$$

or equivalently

$$\frac{B \text{ (1/shipment size)}}{2}$$

Since ordering and handling cost data are extremely difficult to collect, the assumption has been made that this cost factor can be approximated by a constant cost per order. Even with this simplifying assumption, it may be difficult to collect the data required to determine the value of the constant for each firm and commodity being studied. If the constant is not included in the cost function, then the value of the coefficient will reflect the average cost per order for all firms and commodities. If  $\beta$  represents the cost per order, then  $\beta$  multiplied by the frequency of orders will yield an estimate of the annual cost of ordering and handling. Dividing this quantity by the annual usage will produce a per unit cost, as shown in the expression above. Note that this expression may be rewritten in an equivalent form by substituting shipment size for the quotient of the annual usage and the frequency of ordering. If the ordering and handling costs per unit are estimated in this manner, then they will vary with the choice of shipment size, but they will be independent of the choice of a supplier and a mode.

Per unit capital carrying cost  
incurred before receiving the order

$B$  (mean wait time at the origin + mean travel time)

$^3$

\*(FOB price at origin  $i$ )

From the time an order is filled until the time of delivery, the items in the order are unproductive. Hence, there is an opportunity cost of capital associated with a shipment. The key factor which determines the opportunity cost is the rate of return on the investment in the commodities being purchased. However, gathering data on these rates is rather difficult. A simple solution to this problem is to let the coefficient  $\beta$  represent the rate of return. Dividing  $B$  by  $B$  will produce an estimate of the rate in dollars per dollar-day.

It should be noted that the carrying cost is a function of the choices of supplier, shipment size and mode. The mean travel time is directly affected by the choices of supplier and mode. Shipment size may also affect the mean travel time, particularly when the choice of shipment size determines whether an order is carried LTL (LCL) or FTL (FCL).

The waiting time at the origin is also a function of the choices of mode, supplier and shipment size. The influence of the choice of a mode on the waiting time is primarily a function of vehicle availability. Of course vehicle availability is also a function of the choice of supplier (i.e. origin) and shipment size. However the principal impact of the choices of supplier

and shipment size on the waiting time is reflected in the method used to fill the order. Some suppliers fill orders immediately by taking goods from their warehouse. Other suppliers fill orders by scheduling the production of the needed items. In many cases the size of the order will determine which of these two methods is used.

#### Loss of value per unit during transit

$$\beta \frac{(\text{wait time at origin } i + \text{mean travel time})}{4} / \text{shelf life} \\ * (\text{FOB price at origin } i)$$

The loss of value during transit is a significant cost factor in the shipment of perishables such as lettuce, and in the shipment of highly seasonal goods such as Christmas trees or magazines. There are a number of methods available for the calculation of this cost. In the equation presented above, the loss of value is assumed to be proportional to the percentage of the total life of the good which has expired by the time the order arrives at the destination. The constant of proportionality or depreciation rate is reflected in the coefficient  $\beta$ . Dividing  $\beta$  by  $\beta$  will produce an estimate of the rate which is appropriate when the cost is measured in dollars.

As described in the discussion of in-transit carrying costs, waiting time at the origin and mean travel time are functions of

the choice of a supplier, a mode, and a shipment size. Shelf life may also depend on the choice of a supplier, while the FOB price depends on both the supplier and the shipment size [28]. Thus it can be seen that the loss of value is a function of all three choices. However, this cost term should be set equal to zero if the commodity being shipped has an indefinitely large shelf life.

#### Packaging cost, loss and damage per unit

$$B \text{ (packaging cost + loss + damage)}$$

5

In many respects the cost of packaging is complementary to the amount of loss and damage incurred during shipment. The better the packaging, the smaller the costs of loss and damage, and vice versa. Thus it is reasonable to combine all three cost factors into one term of the model. If the packaging cost, loss and damage are all expressed in terms of dollars per unit received, then the coefficient is simply a scaling factor.

It should be recognized that the cost of packaging, loss and damage varies widely between modes. This cost is also sensitive to the choice of a supplier because loss and damage are functions of the length of haul and the region in which the shipment originates. Furthermore, loss and damage are a function of the

-----

28. If the commodity has a non-zero scrap value, then the FOB price minus the scrap value should be substituted for the FOB price in the cost equation.

amount of handling a shipment receives. Therefore the choice of shipment size will also affect the magnitude of this term.

Under some circumstances, the cost of packaging is included in the FOB price, in which case only the loss and damage costs need be considered. It is also common for the cost of loss and damage to be paid by the supplier or the carrier. In this situation, the loss and damage terms should reflect the interest on any of the buyer's money which is tied up while the claim is being investigated. In addition, the receipt of a damaged shipment may cause costly disruptions in the buyer's inventory control system. It is usually impossible to collect data describing this kind of phenomenon. However, the effect of disruptions will be reflected in the estimated value of  $\beta$ .

5

#### Transportation charges per unit

B (tariff + special charges)

6

The tariff is a function of the mode, shipment size, origin and destination. Hence this cost term is sensitive to all three of the choices being modeled. The special handling factor is included to represent any accessorial charges, in-transit processing charges or refrigeration charges. If both the tariff and the special handling charges are expressed in dollars per unit, then the coefficient serves only as a scaling factor.

Capital carrying cost per unit  
incurred after receiving the order

$$B \left( 0.5 * \text{shipment size} / \text{annual usage} \right)$$

<sub>7</sub>

\*(FOB price at origin i)

This term reflects the opportunity cost of capital tied up in the purchaser's stockpile at the destination, except for the safety stock. Note that any given item may be the first or last item to be used from a particular order. On the average, an item is held in stock one-half of the time between orders before it is used. Therefore, one-half multiplied by the length of time between orders, multiplied by the price per item yields the average number of dollar-days of inventory carrying required for each item. As in the case of in-transit carrying cost, the coefficient is used to represent the interest rate. Dividing  $\beta$  by  $B$  will produce an estimate of the interest rate which can be used to express the carrying cost in terms of dollars.

It should be noted that the in-warehouse capital carrying cost is a function of the choice of a shipment size and a supplier. The choice of a mode will have no major impact on this term.

Safety stock carrying cost and stockout cost per unit

$B$  (safety stock carrying cost + stockout cost)

8

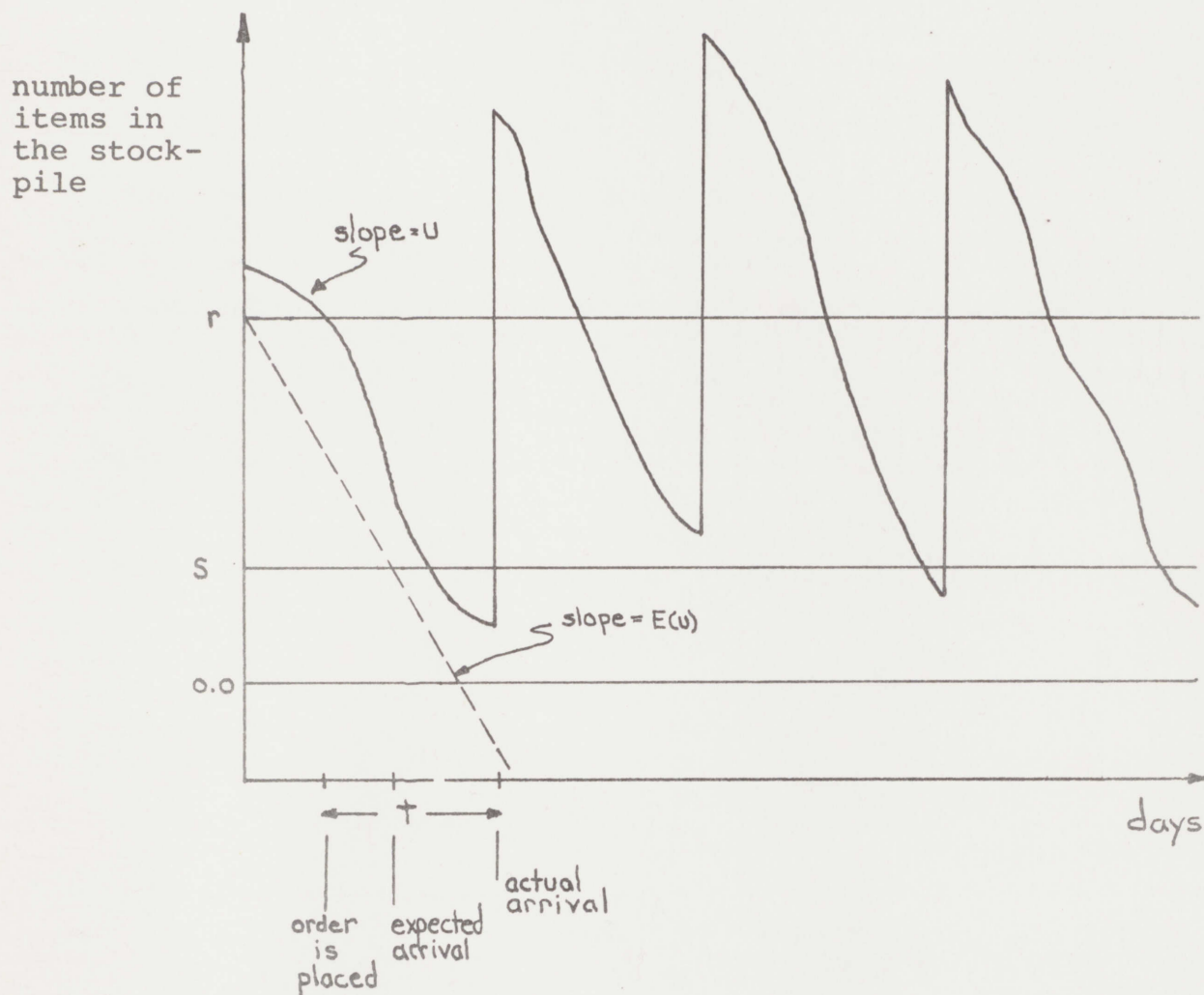
The safety stock carrying cost and the stockout cost result from a rather complex interaction of level of service attributes, receiver attributes and inventory decisions. The following paragraphs address the development of a methodology for analyzing and measuring these two closely related cost factors.

A typical trigger point inventory control system operates in the manner illustrated in Figure 1. Whenever the stock drops below the reorder point  $r$ , an order is sent out. A stockout will occur if more than  $r$  items are needed before the shipment arrives. The number of units needed during the reordering period is a function of the number of days until the shipment arrives and the usage rate on each of these days.

The time which elapses from the placement of an order until the receipt of the shipment varies from order to order. These fluctuations are due to the variation in the time used by the supplier to fill the order, and the unreliability in the carrier's operations. A convenient way of characterizing the situation is with a probability distribution, as shown in Figure 2. The symbol  $P_T(t)$  represents the probability of a shipment arriving  $t$  days after the order is sent out. It should be noted that  $P_T(t)$  will depend on the choice of a supplier, a mode, and a shipment size.

Figure 1

The Reordering Process  
When a Trigger Point System is Employed



$r$  = reorder point

$t$  = time from when an order is placed until it arrives

$u$  = daily usage of the commodity being ordered

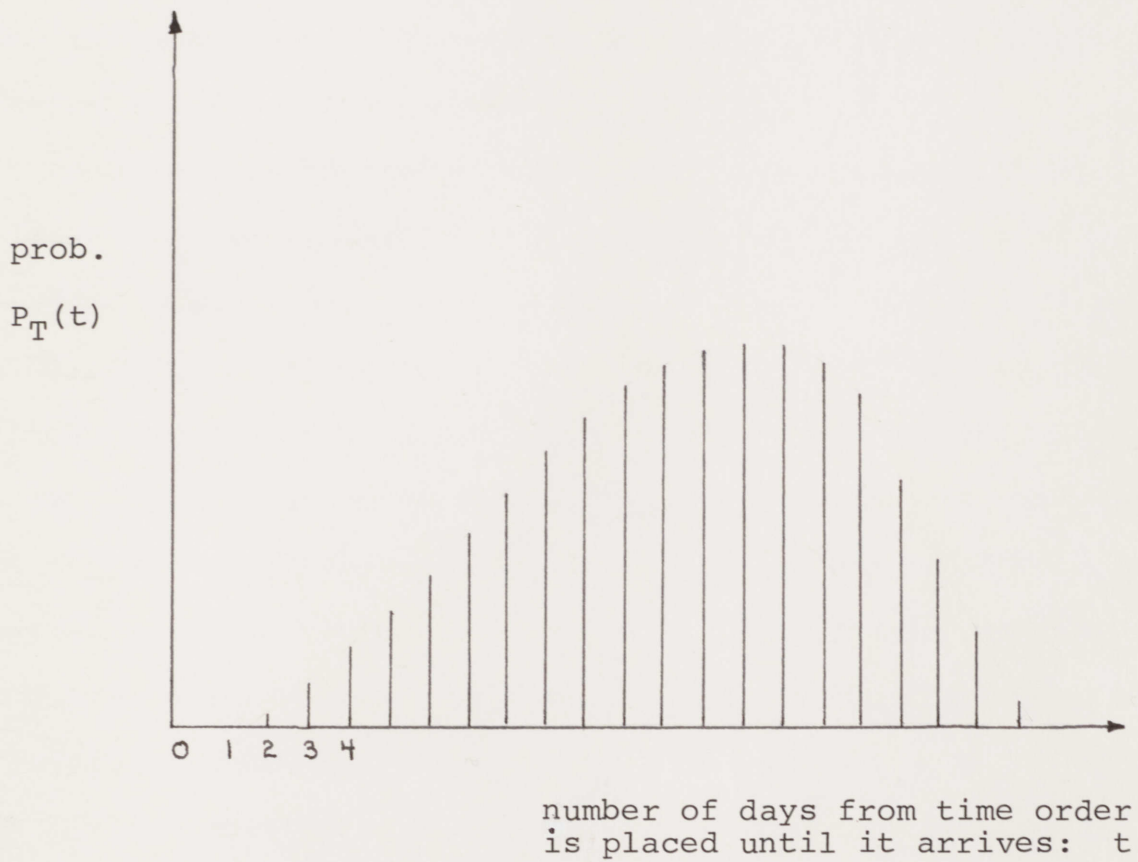
$s$  = safety stock =  $r - (E(t) \cdot E(u))$

$n$  = number of items in stock



Figure 2

Probability Distribution for the  
Procurement Time for an Input

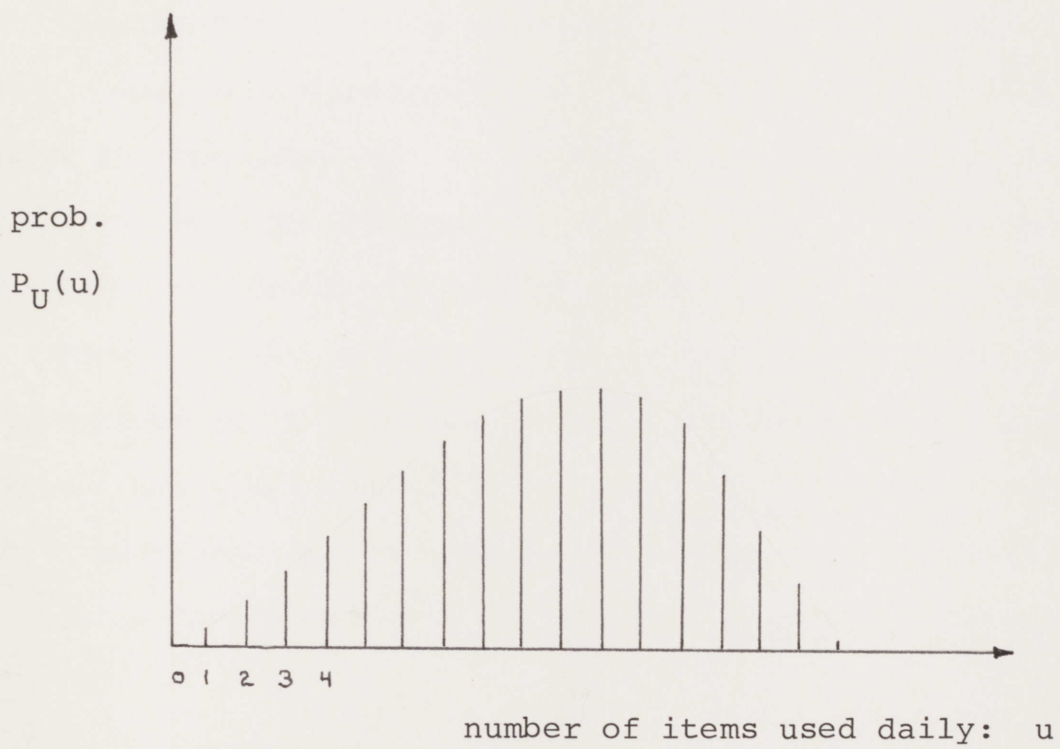


The second random variable in the inventory process is the daily usage rate of the commodity being ordered. This rate varies because of fluctuations in the demand for products, machine breakdowns, absenteeism, etc. These variations can be represented by a probability distribution of the type shown in Figure 3. The symbol  $P(u)$  stands for the probability of using  $u$  items on any given day. Note that this probability is not a function of any of the transportation decisions.

During a reordering period, the expected usage of the input is equal to the product of the expected daily use rate and the expected procurement time because these two variables are statistically independent. The size of the safety stock is defined as the reorder point minus the expected usage. Given the distributions shown in Figures 2 and 3, the safety stock carrying cost can be computed very easily if the reorder point is known. However, data on reorder points are usually not available for the purposes of a demand modeling study. Therefore it is useful to develop a method of calculating the reorder point in terms of more readily available variables.

The safety stock carrying cost and the cost of stockouts are counterbalancing factors. The compromise between the two is established by the stockout risk chosen by the plant manager. For a given supplier, mode and shipment size combination, the stockout risk dictates the minimum reorder point. It is also responsible for the distribution of the sizes of stockouts, in

Figure 3  
Probability Distribution for the  
Daily Usage of an Input



addition to the frequency of stockouts. However, the mathematical relationship between the risk of stockout and the two cost factors is very complex. In a recent paper, Roberts (1975) has developed a set of equations which are relevant to this problem. These equations are similar to the birth and death equations used in queuing theory. Roberts argues that if there are  $(r-i)$  items in stock on a given day, then the probability of having  $n$  items in stock on the following day is equal to  $P(r-i-n)$ , the probability of using  $(r-i-n)$  items in a day. Thus, the probability of having  $n$  items in stock on any given day is equal to the probability of having  $(r-i)$  items in stock on the preceding day, multiplied by the probability of using  $(r-i-n)$  items in one day, summed over all feasible values of  $(r-i)$ . The situation is complicated slightly when the probability of a shipment arriving is also considered. In the general case, the probability of having  $n$  items in stock on day  $t$  is a function of the probability that the shipment will not arrive on or before day  $t$ , as shown in the following equations:

$$P_N(n|T=t, \text{no arrival}) = \sum_{i=0}^{r-n} P_N(r-i|T=t-1, \text{no arrival}) * P_U(r-i-n)$$

$$P_N(n, \text{no arrival}|T=t) = P_N(n|T=t, \text{no arrival}) * (1 - \sum_{z=1}^t P_T(z))$$

where  $P_N(n|T=t)$  = probability of having  $n$  items in stock  $t$  days after an order has been sent out

$P_U(u)$  = probability of using  $u$  items in one day

$P_T(t)$  = probability of a shipment arriving  $t$  days after an order has been sent out

$r$  = reorder point (i.e. the number of items in stock on the day the order is sent out)

$i$  = a dummy counter which ranges over all feasible levels of the inventory

These formulas can be used in a computer program to calculate the reorder point and the distribution of stockout sizes which are associated with a given risk of stockout. Although they are useful, it is obvious that these formulas are too complicated to be used directly in a disaggregate demand model. A method is needed for representing the essential functional relationships in a simpler mathematical form.

Roberts' formulas can be solved for various stockout risks and the locus of solutions can be plotted as shown in Figures 4 and 5. The resulting curves are surprisingly smooth. In Figure 4, the relationship between the size of the safety stock and the probability of stocking out can be approximated by the following expression:

$$s = a_0 \gamma^{-a_1} \quad (1)$$

where  $s$  = size of the safety stock

$\gamma$  = risk of stockout

and  $a_0, a_1$  are parameters to be estimated

Note that equation (1) can be fitted to the curve shown in Figure 4 with regression analysis. Once equation (1) has been estimated, it can be used to express the safety stock carrying cost as follows:

safety stock carrying cost per unit =

$$B \left( a_0 \gamma^{-a_1} \right) * \text{FOB price} / \text{annual usage} \quad (2)$$

As in the carrying cost terms which were discussed earlier in the chapter, the coefficient can be interpreted as the interest rate on capital tied up in the safety stock.

A similar approach can be used to calculate the stockout

Figure 4

The Relationship between the Reorder Point,  
the Safety Stock and the Probability of Stocking Out

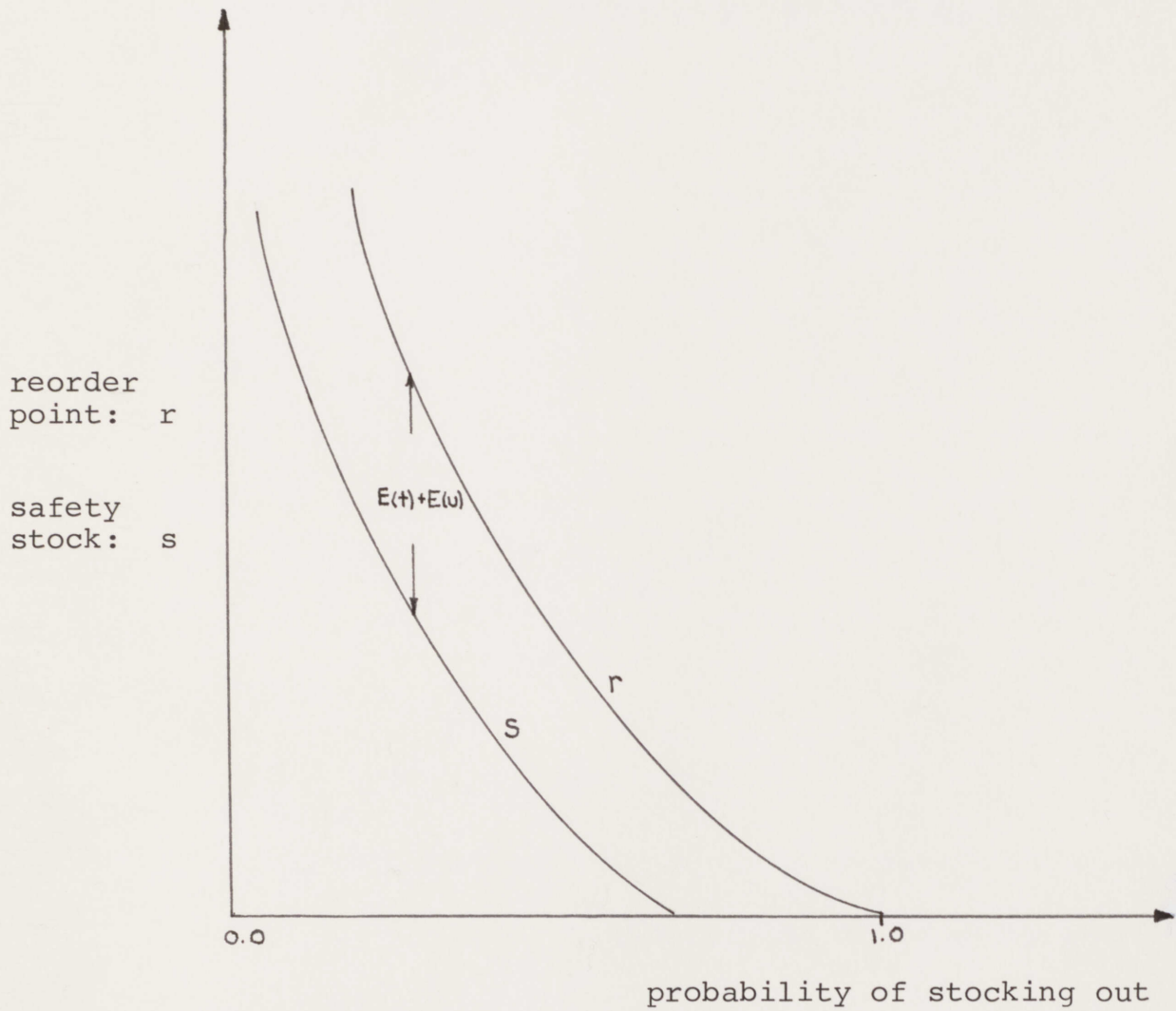
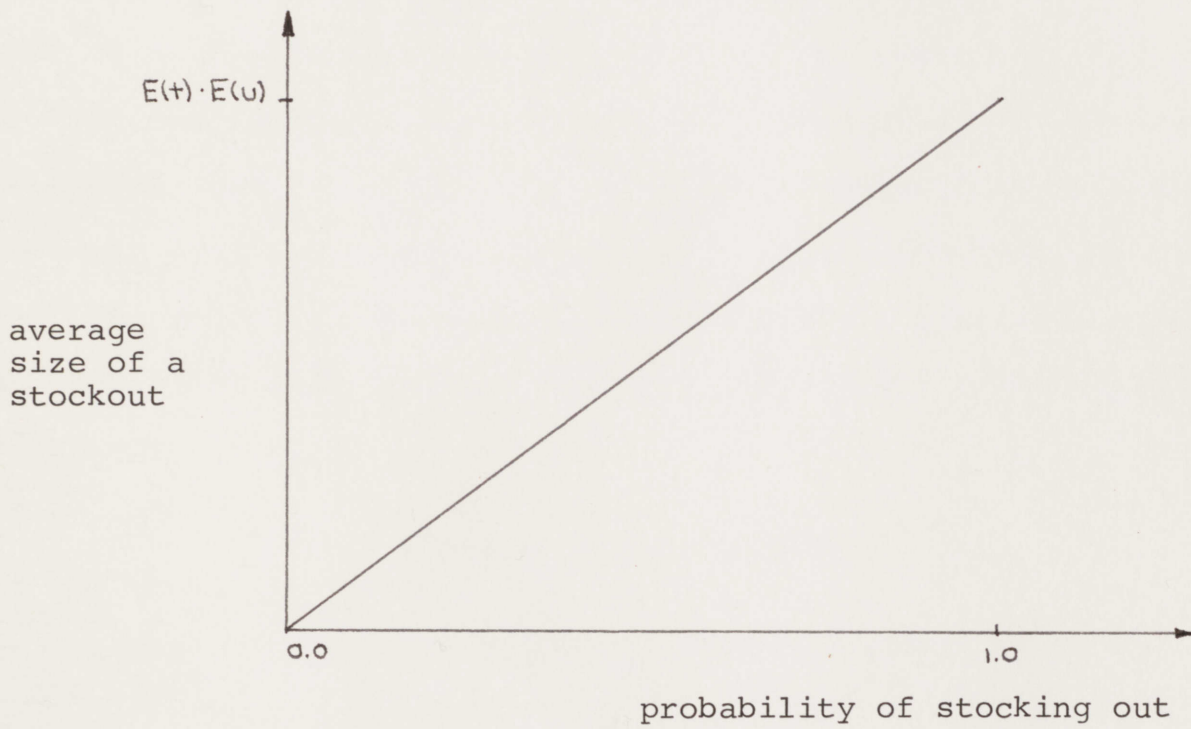


Figure 5

The Relationship between the Average Size of a Stockout and the Probability of Stocking Out





cost. The relationship between the average shortage in a stockout and the stockout risk can be approximated by a straight line, as follows:

$$\text{average shortage} = a \frac{Y}{2} \quad (3)$$

This equation can be fitted to the curve shown in Figure 5 with regression. But in order to use equation (3) to calculate the cost of a stockout, it is necessary to postulate a relationship between the shortage and the cost of a stockout. If no data is available, then a simple but plausible assumption is that the cost of stocking out is proportional to the shortage of inputs and the value of the good being produced. Thus, the cost of stockouts per unit of input could be represented by the following equation:

$$\text{stockout cost per unit} =$$

$$\frac{B}{10} \left( a \frac{Y}{2} \right) * (\text{value of the product}) / (\text{shipment size}) \quad (4)$$

Note that division by the shipment size is necessary to distribute the stockout cost per order over all items in the shipment.

If data can be obtained on the risk of stockout chosen by each firm being studied, then equations (2) and (4) can be used

directly in the demand model. However this kind of information is not generally available and it is difficult to collect. Therefore it is desirable to define the safety stock carrying cost and stockout cost terms in a manner which allows the level of risk,  $\gamma$  to be estimated as a function of the variables already introduced into the model and some set of estimable coefficients.

Adding together the safety stock carrying cost given in equation (2) and the stockout cost given in equation (4) produces the following expression:

$$c(Y) = B_9 (a_0 Y)^{-a_1} (p_k / U) + B_{10} (a_1 Y) (p_o / q) \quad (5)$$

where  $c$  = all logistics costs depending on  $Y$

$Y$  = risk of stocking out of commodity  $k$

$p_k$  = FOB price of commodity  $k$  at origin  $i$

$p_o$  = price or value of the output being manufactured from commodity  $k$

$U$  = annual usage of commodity  $k$

$q$  = shipment size for orders of commodity  $k$

and  $a_0$ ,  $a_1$ ,  $a_2$  are parameters defined as before

Taking the first derivative of equation (5) with respect to  $Y$  and solving for the minimum cost value of  $Y$  produces:

$$Y^* = (B_9 / B_{10})^{1/(a_1+1)} (a_0 a_1 / a_2)^{1/(a_1+1)} (q / p_o)^{1/(a_1+1)} \quad (6)$$

$$* (p_k / U)^{1/(a_1+1)}$$

where  $Y^*$  = the optimal risk of stockout

Note that the optimal risk of stockout depends on the parameters  $a_0$ ,  $a_1$ , and  $a_2$  which are used to represent the effects of the variability in the daily use rate and the procurement time. This implies that  $\gamma^*$  varies as a function of the mode and supplier decisions. The shipment size decision may or may not affect  $a_0$ ,  $a_1$ , and  $a_2$ . However, the explicit presence of shipment size in equation (6) makes it clear that  $\gamma^*$  will depend on this choice also.

The relationship between  $\gamma^*$  and the other variables in equation (6) seems intuitively reasonable. As the value of the good being manufactured increases, the cost of stockouts should increase, and therefore the optimal stockout risk should fall. Since  $a_0$ ,  $a_1$ , and  $a_2$  are greater than zero, equation (6) indicates that  $\gamma^*$  will vary with  $p$  in the expected manner. Also, as the shipment size rises, low stock situations occur less frequently and the safety stock is used less often. Therefore, the size of the safety stock is decreased and the risk of a stockout on any one order increases. This behavior is also captured in equation (6).

Equation (6) can be substituted into equation (5) to produce the following expression for the safety stock carrying cost and stockout cost:

$$c^* = \beta_{11} (x_1 + x_2) \quad (7)$$

where:

$$\beta_{11} = \beta_9 \frac{1/(a_1+1)}{\beta_{10} \frac{a_1/(a_1+1)}$$

$$x_1 = a_0 \left( \frac{a_0 a_1}{a_1} \right)^{-a_1/(a_1+1)} \left( \frac{q}{p_0} \right)^{-a_1/(a_1+1)} \left( \frac{p_k}{U} \right)^{1/(a_1+1)}$$

$$x_2 = a_2 \left( \frac{a_2 a_1}{a_1} \right)^{1/(a_1+1)} \left( \frac{q}{p_0} \right)^{-a_1/(a_1+1)} \left( \frac{p_k}{U} \right)^{1/(a_1+1)}$$

Both  $x_1$  and  $x_2$  can be computed from the data that is assumed to be available. Thus equation (7) can be used in a disaggregate freight demand model to represent the safety stock carrying cost and the cost of stockouts associated with each supplier, mode, and shipment size alternative. If the daily use rate distribution for the commodity is not available, but the procurement time distribution and the expected usage are available, then the methodology developed above can still be used, as will be shown in Chapter 8.

It is important to understand that the use of the optimal stockout risk does not require the assumption that plant managers behave optimally. The coefficient  $\beta_{11}$  is estimated on the basis of the observed behavior of buyers, whether optimal or otherwise. If the estimated value of  $\beta_{11}$  is used in conjunction with equations (6) and (7) to solve for  $\gamma$ , the resulting value will estimate the average chosen risk of stockout, regardless of its nonoptimality. In a sense,  $\beta_{11}$  is being used to represent not only the stockout cost and interest rate, but also the deviation from optimal behavior. Since there are no constraints on the size of this deviation, equation (7) is a perfectly general representation of the safety stock carrying cost and cost of stockouts.

#### Summary of the Specification

If the set of nine composite variables developed in this chapter are used in a disaggregate joint choice model, the resulting model should have a number of desirable characteristics. First, all of the purchasing and logistics costs introduced in Chapter 2 have been included in the specification. This makes it possible to apply the model to many different shipping situations. Secondly, the use of logistics cost principals makes the model specification more intuitively reasonable, and it makes the coefficients easier to interpret.

Thirdly, this specification encompasses a wide range of policy sensitive transport level of service variables. This would make an estimated model useful in the analysis of many current freight transportation issues. And fourthly, the data requirements for estimation and forecasting are not unreasonable. It is true that the data base would have to be considerably more detailed than any of those currently available. However the data collection effort need not be a great deal more complicated than the surveys presently conducted by the Bureau of the Census and the Department of Transportation. Even if all of the data cannot be collected, the specification presented in this chapter establishes a sound theoretical basis for further simplification.

#### Aspects of the Decision Making Process Which Require Special Attention

The specification developed in the preceding section is applicable to a wide range of situations. Nevertheless, it may require some adjustment in response to certain aspects of the decision making process. Modifications to the model might include the alteration of some of the independent variables, the addition or deletion of variables, and the stratification of the data base.

There are five aspects of freight demand which require special attention. One of these is the type of stockout. In

manufacturing, a stockout may halt production or it may force a conversion to another production process. On the other hand, in wholesaling and retailing a stockout may lead to the loss of sales or backorders.

A second element requiring special treatment is the type of inventory control plan. It has been assumed in this chapter that the decision-maker uses a single item trigger point reorder policy. Other possibilities such as single item, periodic reorder and multi-item trigger point plans would require the re-specification of some terms in the cost function.

A third item of interest is the order frequency. It has been assumed that orders are placed fairly frequently. However, some expensive items which are ordered very infrequently are handled in a different manner. A special version of the model may be needed in these cases.

A fourth factor requiring attention is seasonality. The costs discussed above reflect mostly variable costs. However, shipment decisions also involve fixed costs. The seasonal utilization of transport facilities is relevant to the amortization of fixed costs. In addition, the seasonality of the demand for the item being produced is a determining factor in the rate of loss of value of the inputs which are on order.

A final item of special interest is the location of the decision-maker. It has been assumed that the party placing the order makes all shipment decisions. In fact, the supplier at the



origin may participate in the decision making. Special cases of this type require a re-specification of the model.

Thus it can be seen that no single model can address the entire spectrum of freight shipment decisions. However the specification presented in this chapter could provide a very flexible policy analysis tool. The primary obstacles to further development are the theoretical problems with the functional form of the model, and the lack of disaggregate data.

## Chapter 7

## Preparation of a Data Base

As has been indicated in the previous chapters, one of the more imposing constraints against the development of a workable disaggregate freight demand model has been the general unavailability of detailed data. This was no less of a problem in the research presented in this thesis than it has been in past modeling projects. Certainly the estimation of a disaggregate joint choice model of the type described in Chapter 6 will require the collection of a disaggregate data set, with special emphasis put on the collection of shipper/receiver attributes and market attributes.

Unfortunately, initial efforts aimed at collecting a disaggregate data set for use in this thesis research were only partially successful. It became apparent that developing an original data base containing a variety of firms, commodities and modes might require a year or more of work on the part of the author. Since this amount of time was not available, the decision was made to use data from published sources. This has made it difficult to study the choice of a supplier (i.e the origin of a shipment) because data on commodity prices at various points of supply is not publicly available. Thus the data

preparation phase of this research has been directed toward the use of published data to construct a quasi-disaggregate data set that could be used to estimate a model of the mode and shipment size decisions. Of course, the results from previous modeling research tend to indicate that the published data are inadequate in many respects. However, in this study a great effort was made to extract the maximum amount of useful information from the available material. The resulting data set is more enriched and detailed than the data sets constructed by most other researchers who have used the Census of Transportation and similar sources.

#### The Use of Commodity Flow Data

In preparing a data set for the estimation of a disaggregate model, the principal function of commodity flow data is to identify the chosen method of shipment. For this purpose, no better source of information could be found than the O-D File 1 computer tape which was prepared as part of the 1967 Census of Transportation. Each record on this tape includes the number of tons and ton-miles of a given commodity which were sent in a given shipment size category, by a given mode, from a Production Area to a Market Area. The commodities are described by 2, 3, 4, and 5 digit STCC codes. The list of modes includes rail, common carrier truck, private truck, air and water. Each Production Area and each Market Area is composed of a small group of

SMSAs [29]. The twenty shipment size categories are defined in Appendix III. In most cases, a shipment size category spans several thousand pounds. For purposes of the work described herein, a mean shipment size has been assumed in each category. These assumed mean shipment sizes are also listed in Appendix III.

Each record in the C-D File 1 tape contains one other very useful piece of information. This is the number of observations in the Census Bureau waybill sample which have the characteristics associated with that record. In other words, for each origin-destination-commodity-mode-shipment size record on the tape, there is a count of the number of shipments of that type which were found in the 1.4 million waybills collected in the 1967 Census of Transportation survey of the Major Industrial Sector. As described in Chapter 4, this survey is composed of a probability sample of all shipments of manufactured goods which originated from plants with more than twenty employees. If this survey had been available, then it could have been used to estimate disaggregate freight demand models. Since this survey is not available, an alternative approach is to reconstruct the sample using the counts given on the tape, the assumed mean shipment size in each size category and the other descriptions included in each record. For example, suppose that a record on

---

29. A listing of the SMSAs in each Production Area and Market Area is included in Appendix II.

the tape indicates that the Census Bureau's waybill sample contains three waybills for five - ten ton shipments of sulfuric acid which were sent from St. Louis to Baltimore by rail. From this information we could synthesize three individual records of 7.5 ton shipments of acid sent by rail from St. Louis to Baltimore. This is the basic technique that was used to generate a disaggregate sample of the chosen methods of shipment for use in this research.

From the list of twenty-five Production Areas and fifty-five Market Areas used in the 1967 Census of Transportation, eighteen O-D pairs were chosen for further study. This selection was based largely on the availability of the rail travel time data which will be discussed later in this chapter. However an effort was made to include a variety of commodity flows and lengths of haul. The O-D pairs listed by number and the name of the largest SMSA in each Area are shown in Table 1.

For each O-D pair, all records of commodity flows which were described by 5-digit STCC were skimmed from the 1967 Census of Transportation O-D File 1 tape. Only 5-digit STCC were used because a maximum amount of commodity detail is required in order to determine the attributes of the good being shipped. It should be noted that this sample has a bias toward commodities which are manufactured by a large number of firms. This is the case because the Census Bureau will disclose O-D flows at the 5-digit STCC level only if there are five or more firms manufacturing the

Table 1

## Origins and Destinations Included in the Data Base

<u>Production Area (origin)</u>	<u>Market Area (destination)</u>
Buffalo 10	Boston 1 Philadelphia 5 Cincinnati 14
Cleveland 11	Cincinnati 14 Memphis 39
Detroit 13	Buffalo 10 St. Louis 18
Cincinnati 14	Boston 1 Philadelphia 5 Cleveland 11 Atlanta 19 Birmingham 42 New Orleans 45
St. Louis 18	Boston 1 Philadelphia 5 Pittsburgh 12 Cincinnati 14 Atlanta 19 Louisville 37

commodity in the origin area.

The skimming of data from the tape yielded 650 different origin-destination-commodity-shipment size records. In the manner described above, each record was replicated several times. This resulted in a data set containing 1430 shipments, of which 1300 went by truck and 130 went by rail. A cross tabulation of the data by mode, length of haul and shipment size is given in Tables 2 and 3. Having developed a quasi-disaggregate data set containing the chosen mode and shipment size for these 1430 shipments, the next step is to enrich each observation with receiver attributes, level of service attributes and commodity attributes.

Table 2

## Cross Tabulation of Rail Shipments by Size and Distance

		Shipment Size (tons)				
		under 1	1-5	5-20	20-40	over 40
Length of haul (miles)	under 100	0	0	0	0	0
		0	0	0	0	0
	100-500	0	3	8	13	15
		0	4	18	15	28
	500-1000	0	2	9	9	1
		0	3	17	26	1
	over 1000	0	0	1	5	0
		0	0	1	17	0

Note: The upper entry represents the number of flows reported at the 5-digit STCC level in the 1967 Census of Transportation O-D File 1 Tape for the O-D pairs shown in Table 1.

The lower entry represents the number of disaggregate observations generated from the aggregate Census data.



Table 3

## Cross Tabulation of Truck Shipment by Size and Distance

		Shipment Size (tons)				
		under 1	1-5	5-20	20-40	Over 40
Length of haul (miles)	under 100	0	0	0	0	0
		0	0	0	0	0
	100-500	149	140	92	33	2
		321	355	224	93	5
	500-1000	82	49	25	8	0
		142	106	35	14	0
	over 1000	2	2	0	0	0
		3	2	0	0	0

Note: The upper entry represents the number of flows reported at the 5-digit STCC level in the 1967 Census of Transportation O-D File 1 Tape for the O-D pairs shown in Table 1.

The lower entry represents the number of disaggregate observations generated from the aggregate Census data.

### The Development of Receiver Attributes

One vital piece of information which cannot be recovered from the Census of Transportation data is the name of the firm which received a given shipment. This makes it very difficult to develop a description of the receiver or decision-maker. However a rather elaborate procedure was developed to estimate the mean usage rate of the average purchaser of any given commodity in any of the Market Areas (i.e. destinations).

The estimation of usage rates is difficult because most of the available data describing the manufacturing sector is oriented toward the output side. There is very little information available on the value of inputs used by industries in different cities. Therefore, it is necessary to infer the amount of input consumed by an industry based on the volume of output of that industry. The key information which makes these inferences possible are the technical coefficients used in the Leontief Input - Output model. Each technical coefficient in this model represents the value of a given input required to produce one dollar's worth of a given output. Thus, if the value of output is known for each industry in a Market Area, then the consumption of each commodity by each industry can be computed.

The most straight-forward method of estimating the production volume of each industry in a Market Area is to use the data contained in the Census of Manufacturers. Unfortunately, the Census of Manufacturers lists the output of industries by

SMSA, not by Market Area. Therefore the production of each industry must be summed across all SMSAs in each Market Area. This task is rather large because some Market Areas contain more than five SMSAs, and most SMSAs contain many industries. Furthermore, the 1967 Census of Manufacturers data is not available on computer tape, and therefore the data would have to be processed by hand. Due to the time constraints on this research, a faster method of computing production levels had to be devised.

The 1967 County Business Patterns data is readily available on computer tape. This reference contains information on the number of firms in each of eight employment size groups, for each industry, in each county in the country. By assuming a mean number of employees per firm in each of the eight employment size groups, it is possible to use this data to compute the total employment in each industry, in each county [30]. Deriving the employment in each industry, in each Market Area is simply a matter of using the computer to sum up data from the counties in each Market Area.

The real problem arises in converting the employment per industry into the volume of output per industry. Due to the time constraints, the simplest and crudest approximation had to be used. The 1967 Census of Manufacturers data was used to derive a

---

30. Total employment data is given for some, but not all industries listed in the County Business Patterns.

national average productivity per worker in each industry. These productivities were then multiplied by the employment figures to produce an estimated value of total output for each industry, in each Market Area.

The inputs consumed by an industry can be estimated by multiplying the value of their output by the vector of technical coefficients corresponding to that industry. The vectors of technical coefficients used in this study were computed from a large 1967 national input-output table which was prepared by the Bureau of Economic Analysis [31]. These coefficients were used in conjunction with the value of output data to derive the total consumption of each commodity by each industry in the Market Areas of interest. As a final step, data from the County Business Patterns were used to compute the number of firms per industry, which was used in turn to compute the average consumption per consuming firm, for each commodity in each Market Area.

The procedure described above makes it possible to compute an average usage rate for the average size purchaser of any commodity in any city. Obviously, this derived data falls short of the quality and quantity of data that should be collected in a good disaggregate survey. Given any one shipment synthesized from the data contained in the Census tape, the average usage

---

31. Interindustry Economic Division, U.S. Bureau of Economic Analysis, 1967 National Input-Output Table, 484 Industry Level, tape no. 1106.

rate of the average size consumer may be a poor approximation of the average consumption rate of the firm which actually received the shipment. Furthermore, the average usage rate does not reflect the variability in the daily usage rate of the firm. Thus, the average usage rate of the average consumer may have little to do with the conditions which led to the shipment decision that was recorded on the Census tape.

The importance of the missing data on usage rates cannot be overstated. The Census of Transportation indicates that in almost all cases, several modes and shipment sizes are used to transport a commodity between a pair of cities. In theory, if all purchasers of a commodity were identical, then only one mode and shipment size would be utilized for the carriage of a commodity between a given O-D pair. Conversely, the variety of modes and shipment sizes which are actually utilized indicates that there are a variety of firms ordering the commodity.

It should be noted that the procedure described for the derivation of use rates could be refined. It is possible to derive the entire distribution of average use rates of all consumers. This distribution could be used in several ways. From a theoretical standpoint, the proper approach would be to integrate the disaggregate model over the distribution of use rates before estimation. However, this approach is analytically intractable in many cases. A second approach is to integrate the cost function of the model over the distribution of use rates

before estimation. This approach can always be applied. Of course, the data base needs to be much larger if a distribution is used in place of disaggregate data. Due to time constraints, the derivation of use rate distributions and integration techniques could not be pursued in this research project.

Aside from the usage rate data, there are several other pieces of information pertaining to the receiver which would be gathered if a survey was taken, but which are not available for shipments synthesized from the Census tape. The first of these is the type of inventory system that is used. As indicated in the preceding chapter, the specification of the model should be adjusted to the type of inventory system. For lack of data, it has been assumed that all receivers use a single item trigger point system. A second item of missing information is a description of the stockout situation. If these data were available, then it would be possible to estimate separate models for firms which simply lose sales in the event of a stockout, and for firms that must close down assembly lines if a stockout occurs. Since this type of information is not available, the stockout term in the model must be specified in a somewhat vague manner. Finally, a third piece of missing information is the role of the supplier in the decision making process. It has been assumed that the receiver has complete freedom in choosing a mode and shipment size. However suppliers usually have some direct or indirect influence on transport decisions. Unfortunately, this

influence cannot be described with the available data.

### The Development of Commodity Attributes and Level of Service Data

The use of five digit STCC flow data in this study makes it relatively easy to gather commodity attribute data. One key variable is the value per pound. As discussed in Chapter 6, this variable influences the carrying cost. It also affects the cost of loss and damage, and the transport tariff. A second key variable is the density of the commodity being shipped. The primary influence of density is on the transport rate. In the long run, density is also a factor in the size (and cost) of the warehouse needed by the receiver. However in the short run it has been assumed that the size of the receiver's warehouse is fixed at some level large enough to accommodate all shipment size alternatives under consideration. Both value and density data are available in a recent publication titled "A Commodity Attribute File for Use in Freight Transportation Studies" [32]. This reference also contains data on the state of the commodity (e.g. solid, liquid, gas, etc.), the shelf life and special handling requirements. This additional information was not used extensively in this study because almost all commodities in the sample of shipments were solids with indefinitely long shelf

---

32. Samuelson, R. and Roberts, P., "A Commodity Attribute File for Use in Freight Transportation Studies", Report 75-20, Center for Transportation Studies, M.I.T., 1975.

lives and no special handling requirements.

It should be noted that the commodity value data used in this research represents an estimate of the national average FOB factory price of the good. Data on the price of a commodity in various cities or regions is not generally available. However for a mode and shipment size model, the national average price is sufficiently accurate.

Once the commodity attributes have been determined, they can be used in the derivation of level of service attributes. In particular, the value and density data can be used along with distance and shipment size to estimate the transport rate for each alternative. As mentioned in Chapter 3, it is infeasible to use tariff books to look up rates on a large number of shipments. Therefore a simple method of estimating rates is required. Fortunately, this author was able to use results from a fairly extensive study of rate estimation models presently underway at M.I.T. [33]. Although numerous other researchers have developed equations for estimating transport rates, the M.I.T. study has the advantage of using large disaggregate data sets which cover both rail and truck rates [34]. The final results of this research are not yet available; however three preliminary models

---

33. This study of rate models is one part of the research on contract CO-04-50154-00, "Analysis of the Incremental Cost and Trade-Offs Between Energy Efficiency and Physical Distribution in Intercity Freight Markets", sponsored by the Federal Energy Administration.

34. Unfortunately, these disaggregate data sets are proprietary and unavailable for other applications.



have been used to estimate rates for this freight demand modeling research.

The rate models are shown in Table 4. Note that separate models are used to predict full truckload and less than truckload rates. Since the Census of Transportation data does not differentiate between FTL and LTL shipments, a question arises as to which rate model should be applied to each truck-shipment size alternative. A review of the data used to estimate the rate models indicates that the minimum weight required to qualify for a FTL rate varies widely from commodity to commodity. In general, shipments over 20,000 pounds are billed according to the FTL tariff. Experiments with the rate models indicate that for shipments with weights between five and ten tons, the lower of the two rates predicted by the FTL and LTL models is usually a good approximation of the true rate. These results provide a simple set of rules for the use of the truck rate models.

Other level of service attributes which are required for the estimation of freight demand models include the value of loss and damage. In many cases, the cost of loss and damage are ultimately absorbed by the carrier. Undoubtedly, loss and damage disrupt the inventory control process of the purchaser. Since the cost of this disruption is difficult to measure, an alternative procedure is to use the cost of in-transit loss and damage as a proxy for the cost actually incurred by the purchaser

Table 4  
Freight Rate Estimation Models

$$\text{rate (cents/100 lbs.)} = \prod_i x_i^c$$

$\frac{x}{i}$	Rail-FCL	Truck-FTL	Truck-LTL
1. distance (miles)	.549	.226	.305
2. distance/250 if distance > 250		.426	
3. distance/300 if distance > 300	.152		
4. distance/500 if distance > 500			.166
5. weight (lbs.)	-.767	-2.003	-.156
6. weight/6000 if weight > 6000		1.317	
7. weight/40000 if weight > 40000	.355		
8. weight/120000 if weight > 120000	.305		
9. value (\$/lb.)	.153	.069	.023
10. density (lb./ft. <sup>3</sup> )		-.079	-.169
11. gas (1=yes, 0=no)	.308		
12. constant	9.196	21.854	5.451
	2 R = .89	2 R = .85	2 R = .71

or decision-maker.

The principle sources of the loss and damage data used in this research are the quarterly reports on truck L/D which are published by the I.C.C., and the annual reports on rail L/D which are published by the A.A.R. Both of these reports contain the value of claims listed by commodity and cause. Since the model specification does not address specific causes of L/D, the data were aggregated over all causes. Then the claims in each commodity group were normalized by the volume of shipments of that commodity. This resulted in an estimate of the value of claims per ton of shipment of each commodity. Note that loss and damage could also be expressed in terms of dollars per ton-mile. However the dollars per ton statistic is probably more relevant for durable goods because the loss and damage of these commodities occurs mostly in origin and destination terminals.

The loss and damage data derived from the A.A.R. and I.C.C. bulletins has many shortcomings. First, the breakdown of L/D by commodity is not very extensive. This makes it necessary to assume that all 5 digit STCC commodities in a 2 or 3 digit STCC group have the same rate of loss and damage. Also, the lack of L/D data for specific cities makes it necessary to assume that the rate of loss and damage is constant throughout the country. Neither of these assumptions is very realistic. Hence it is questionable whether the L/D data derived for this study is adequate for use in describing the shipment alternatives.

The final type of level of service data which must be collected are travel time distributions. In the case of rail, actual origin yard to destination yard travel time distributions were made available to the author on a confidential basis. The only modification made to these data was the addition of one day of travel time to represent the pickup and delivery time at the origin and destination.

In the case of truck, no good source of travel time distributions could be located. This made it necessary to estimate the mean travel time for each O/D pair. The values used to make these estimates are shown in Table 5. Although this approach represents a very crude estimation procedure, the results were substantiated by a telephone poll of carriers.

Table 5

## Mean Travel Time for Truck Shipments

<u>Travel Time</u>	<u>Maximum Length of Haul</u>	
	<u>LTL</u>	<u>FTL</u>
1 day	300 mi.	400 mi.
2 days	700 mi.	900 mi.
3 days	1100 mi.	1400 mi.

For lack of other information, it has been assumed that truck deliveries are always on time. This is certainly not a realistic assumption, although truck travel times are usually much more reliable than rail.

It should be noted that there are several level of service variables which are not represented in the data base because they cannot be accurately estimated using the techniques currently available. Included in this class of variables are handling costs, packaging costs and special service charges. The effect on the demand model of omitting these variables is difficult to judge, although it could be significant in some cases.

### Summary

The objective of the data preparation task was to create a quasi-disaggregate data set for the estimation of a model of the mode and shipment size decisions. As shown in this chapter, an adequate sample of shipment decisions can be constructed from published sources. But, the complete set of attributes of the decision-maker and the attributes of the shipment alternatives cannot be estimated with great accuracy. This is due in large part to the fact that the identity of the shipping and receiving firms cannot be determined. It appears that the only real solution to these problems is to collect a disaggregate data set from field interviews with shippers and receivers.

## Chapter 8

### Estimation Results

The model estimation phase of this research has been focused on the mode and shipment size decisions. The majority of the effort has gone into modeling the choice of a mode for a given shipment size. The results from several specifications of a conditional mode choice model will be discussed in this chapter. Also, the results from some initial experimentation with a model of the joint choice of mode and shipment size will be presented. However, the state of the art in modeling joint choices in freight transportation is still rather primitive. Thus, the empirical results from the work with the joint choice model must be considered as only preliminary.

#### The Conditional Mode Choice Model

Since it has been argued that a mode and a shipment size are selected jointly, the rationale for the estimation of a conditional mode choice model may not be clear. The primary reason why this type of model was investigated is because it presents no immediate theoretical problems. The well known disaggregate logit model can be applied to the conditional choice of a mode without modification to the model structure or the

computer software used to estimate the model. In contrast, none of the disaggregate models presently in use can be applied to the joint choice of a mode and a shipment size without violating the basic assumptions used to derive the models. As discussed in Chapter 5, a joint choice model for qualitative and continuous alternatives has not been used in previous studies. Additional research is needed into the theoretical development of the required model.

A second reason why the conditional mode choice model was studied was to gain experience with the data base. Previous researchers had never prepared the Census of Transportation data in the manner described in Chapter 7. Nor had they used as elaborate techniques to estimate transport tariffs or receiver attributes. Hence, the strong and weak points of the methods used to synthesize the data base were not known when this research began. Therefore, the initial estimation work was designed (in part) to uncover problems with the data. It was decided that the data should be tested first with a conditional mode choice model, before moving on to a joint choice model that would be more demanding of the data. Unfortunately, experimentation with the mode choice model brought to the surface many problems with the data base, and consequently a great deal of time was spent on refining the methods used to estimate the independent variables. Among other refinements, the rate models were replaced four times and the use rate calculations were

revised three times in the course of the research.

The four conditional choice mode choice model specifications discussed in the following paragraphs are presented in the order in which they were developed. Although the data base was revised several times during the research, the estimation results for these four specifications have been updated to the latest data base, which was described in Chapter 7.

#### Definition of the Set of Alternatives

As explained in Chapter 7, the Census of Transportation O-D File 1 Tape was used to synthesize a disaggregate sample of shipments. Although the Census data adequately define the chosen method of shipment, they do not include a description of the unchosen alternatives. Thus in setting up a mode choice model, some assumptions had to be made concerning the set of options available to each decision-maker. The lack of data on the quality of service afforded by barge and air transport makes it difficult to include these two modes in the choice set. On the other hand, the methods described in Chapter 7 make it possible to estimate the tariff and travel time for rail, LTL truck, and FTL truck. Therefore the data set used for model estimation was limited to observations in which the chosen mode was either rail or truck. It has been assumed that each truck shipment could have been sent by rail. It has also been assumed that each rail shipment could have been sent by either LTL truck or FTL truck,



depending on the shipment size. The LTL alternative was included in the choice set if the shipment size was less than five tons, and the FTL alternative was included if the shipment size was over ten tons. For shipments with a weight between five and ten tons, the LTL alternative was included if the estimated LTL rate was lower than the estimated FTL rate, otherwise the FTL alternative was included. This use of the estimated rate data may seem rather unusual. However the author's review of some confidential disaggregate shipment data indicates that this method of determining whether a shipment is large enough to qualify for FTL service is more accurate than rules based on only the shipment weight or volume. It should be noted that a serious bias in the estimated values of the model coefficients could result from the use of an independent variable to define the set of relevant alternatives. No problem arises in this particular case because the FTL and LTL alternatives are never available for the same shipment.

### The Naive Specification

Given that the set of alternatives has been defined in the manner described above, the conditional mode choice model requires the specification of three cost functions: one for rail, one for FTL and one for LTL. The main purpose of the model estimation work is to demonstrate the use of logistics management theory in specifying these three functions. However for purposes

of comparison, the conditional mode choice model was tested once with a specification which does not utilize any logistics management principals. This so-called naive specification involves the use of level of service attributes and commodity attributes as individual items in the cost functions. No attempt was made to include any logistics cost variables of the type described in Chapter 6. Furthermore, no receiver attributes have been included. The variables used in this specification are defined in Table 1.

The cost functions used in the naive specification are shown in Table 2. There are several features of this specification which should be noted. First, the level of service variables are used generically. This approach makes it somewhat easier to apply the estimated model to a situation involving a 'new' alternative. On the other hand, the generic use of a variable has the disadvantage of constraining the cost per unit associated with that variable to be the same for all alternatives. Since most previous researchers have used level of service variables generically, the decision was made to use the generic approach in the initial specification.

A second feature of the naive specification is the use of shipment size, value and density as alternative specific variables. These variables cannot be used generically because they do not vary between alternatives. However variables of this type may be included in one or two of the three cost functions.

Table 1

## Definitions of Variables Used in the Naive Specification

<u>Variable</u>	<u>Definition</u>
TT	mean travel time (days)
LD	loss and damage (dollars per pound)
RATE	transport rate (dollars per pound)
RC	(1 for rail, 0 otherwise)
LC	(1 for LTL, 0 otherwise)
RDIST	rail distance (miles)
LDST	highway distance (miles)
O	shipment size (pounds)
VAL	commodity value (dollars per pound)
DEN	commodity density (lbs. per cubic foot)

Table 2

Specification of the Cost Functions for the  
Naive Conditional Mode Choice Model

<u>Coefficient</u>	<u>LTL</u>	<u>Alternative</u> <u>FTL</u>	<u>Rail</u>
B <sub>1</sub>	TT	TT	TT
B <sub>2</sub>	LD	LD	LD
B <sub>3</sub>	RATE	RATE	RATE
B <sub>4</sub>	0	0	RC
B <sub>4</sub>	LC	0	0
B <sub>6</sub>	0	0	RDST
B <sub>7</sub>	LDST	0	0
B <sub>8</sub>	0	0	Q
B <sub>9</sub>	Q	0	0
B <sub>10</sub>	0	0	VAL
B <sub>11</sub>	VAL	0	0
B <sub>12</sub>	0	0	DEN
B <sub>13</sub>	DEN	0	0

When this technique is used, the value of the coefficient reflects the difference between the effect of the variable on the attractiveness of a specific alternative, and the effect of the variable on the attractiveness of all other alternatives. In this case, the alternative specific variables are used in the LTL and rail cost functions, but not in the FTL function. Hence  $\beta_8$  through  $\beta_{13}$  are measuring the influence of shipment size, value and density on the use of LTL and rail relative to the impact of these three variables on the use of FTL. It should be noted that distance has also been used as an alternative specific variable, although its coefficient should be interpreted like the coefficient of a generic variable because distance varies from alternative to alternative.

It is important to understand that alternative specific variables such as value, density and shipment size are undesirable. It is true that they improve the fit of the model by adding information concerning the circumstances surrounding the shipment decision. On the other hand, these variables are a reflection of the degree to which the model must rely on the circumstances, rather than the attributes of the alternatives to explain shipment decisions. The alternative specific variables such as value and density do not explicitly capture the interaction between the circumstances and the attributes of the alternatives. A better approach is to combine the commodity attributes and receiver attributes with the level of service

attributes so that the interactions can be captured explicitly in the policy sensitive variables. This is the primary reason why the naive type of specification was not investigated more extensively in this research.

The results from the estimation of the naive conditional mode split model are shown in Table 3. Since the level of service variables are directly proportional to the cost of an alternative, they should have negative coefficients. Note that the coefficient of the transport rate is positive. The t statistic indicates that this coefficient is not significantly different from zero. However, it cannot be reasonably assumed that rate has no effect on the choice of mode, or that raising the rate would increase the attractiveness of a mode. Thus, it must be concluded that there is a serious error in the specification of the model or in the measurement of the independent variables.

The coefficients of the alternative specific variables are more difficult to interpret. However, the sign of  $\beta_{10}$  is suspicious. If  $\beta_{10}$  is positive, then the model will predict that rail becomes increasingly attractive relative to FTL as the value of the commodity increases. This relationship is not plausible.

In conclusion, the model resulting from the naive specification is unacceptable. It is likely that these results are due in large part to errors in the specification. However

Table 3

## Estimation Results from the Naive Specification

<u>Variable</u>		<u>Coefficient</u>	<u>t_Statistic</u>
1) Mean travel time	TT	-.138	-1.33
2) Loss and damage	LD	-6487.3	-4.70
3) Transport rate	RATE	+.357	+0.76
4) Constant (rail)	RC	-4.14	-6.59
5) Constant (LTL)	LC	-.815	-0.86
6) Distance (rail)	RDST	+.0015	+1.56
7) Distance (LTL)	LDST	-.00027	-0.18
8) Shipment size (rail)	Q	+.00005	+8.16
9) Shipment size (LTL)	Q	-.00016	-2.35
10) Value (rail)	VAL	+1.205	+3.90
11) Value (LTL)	VAL	+6.746	+2.16
12) Density (rail)	DEN	-.0042	-1.64
13) Density (LTL)	DEN	+.034	+2.31

no. of observations = 1430

\*

L (0) = -991.3 (log likelihood for coefficients of zero)

\*

L (B) = -236.3 (log likelihood for estimated coefficients)

2

pseudo r<sup>2</sup> = .76 (explained log likelihood/total log likelihood)

the strong influence of the distance and the commodity value on the transport rates suggests that some of the problems may be caused by errors in the estimation of the rates. Although numerous improvements could have been made to the naive specification, the research was directed instead toward the use of the composite variables which were discussed in Chapter 6.

#### Specifications Based on Logistics Cost Concepts

The primary goal of the model estimation phase of this research is to demonstrate the use of logistics cost concepts in specifying disaggregate freight demand models. Unfortunately a conditional mode choice model is not a very good model for the purpose of this demonstration. Of the eight composite variables developed in Chapter 6, only four are directly relevant to the choice of a mode when the shipment size is a given condition. The purchase cost, order and handling cost, and capital carrying cost incurred after receiving the order are relevant only to the shipment size decision. Furthermore, the loss of value during transit was not considered in this research because all of the commodities in the sample have an indefinite shelf life.

One of the variables which was developed in Chapter 6 and which relates to the choice of a mode is the capital carrying cost incurred before the arrival of the shipment at the destination. It has been assumed that the decision-maker at the destination must pay the capital carrying cost in all cases. It



was necessary to eliminate the waiting time at the origin from this variable because of the lack of data. Thus, this variable was computed by multiplying the average travel time by the value of the commodity being shipped. The resulting quantity will hereafter be referred to as the in-transit capital carrying cost.

A second variable of interest is the cost of packaging, loss, and damage. The cost of packaging had to be ignored in this research because no data was available. The cost of loss and damage on rail and truck shipments can be estimated using the technique described in Chapter 7. However, it is not possible to differentiate between the rate of loss and damage on FTL shipments and the rate on LTL shipments.

The transport charges represent another logistics cost factor which has an influence on the selection of a mode. The rates can be estimated for rail, LTL, and FTL by using the regression equations presented in the previous chapter. It should be noted that inaccuracies in the estimation of the tariffs has caused problems throughout this research. Although the regression equations have been revised four times, they may still be too inaccurate for use in this application. It should also be noted that no attempt has been made to estimate the charge for special services such as refrigeration.

The fourth composite variable that should be considered is the combined cost of carrying a safety stock and stocking out. A rather elaborate method for estimating this variable was

developed in Chapter 6. However, the limitations in the data have made it impossible to fully utilize this technique. In the first place, this cost cannot be estimated for truck shipments because of the lack of travel time distributions for either LTL or FTL service. Thus, the safety stock carrying cost and stockout cost have only been used in reference to the rail mode. Secondly, the mean use rate of the average size purchaser had to be substituted for the use rate distribution of the actual purchaser. And thirdly, in the computation of the cost of stockouts, the value of the commodity being shipped had to be used in place of the price of the commodity being produced. Obviously each of these changes severely undermines the quality of this variable. Hence, the empirical results described in the following paragraphs should be taken as only a preliminary test of this variable as it was defined in Chapter 6.

The variables used in the model specifications based on logistics concepts are shown in Table 4. During the course of the research, many other variables were also tried, including the total O-D flow and the production rate of the average manufacturer of the commodity being shipped. Most of these variables performed poorly, and therefore they have not been included in the models which are presented in this chapter. However, distance, use rate, and shipment size have been included in Table 4 because tests indicate that they significantly improve the fit of the model.

Table 4

Definitions of Variables Used in  
Specifications Based on Logistics Costs

<u>Variable</u>	<u>Definition</u>
ITCC	in-transit capital carrying cost (dollar-days per pound)
LD	loss and damage (dollars per pound)
RATE	transport rate (dollars per pound)
SSSC	safety stock carrying cost and stockout cost (dollars per pound)
RC	(1 for rail, 0 otherwise)
LC	(1 for LTL, 0 otherwise)
RDST	rail distance (miles)
LDST	highway distance (miles)
USE	mean usage rate of the average size firm (pounds per year)
Q	shipment size (pounds)

Table 5

A Logistics Cost Based Specification of the  
Conditional Mode Choice Model

<u>Coefficient</u>	<u>LTL</u>	<u>Alternative</u> <u>FTL</u>	<u>Rail</u>
B 1	ITCC	ITCC	ITCC
B 2	LD	LD	LD
B 3	RATE	RATE	RATE
B 4	0	0	SSSC
B 5	0	0	RC
B 6	LC	0	0
B 7	0	0	USE
B 8	USE	0	0
B 9	0	0	RDST
B 10	LDST	0	0

One of the logistics cost based specifications which has been tested is shown in Table 5. It should be noted that three of the four logistics cost variables are used generically. The safety stock - stockout cost term has been used only in the rail cost function because of the data problems which were discussed above. It should also be noted that in this specification, the logistics cost variables integrate the commodity attributes and receiver attributes with the level of service attributes. Nevertheless, tests have shown that models which utilize only the four logistics variables perform poorly. Therefore it is necessary to add some alternative specific variables. The use rate of the average purchaser was added to the specification to represent the relative cost of operating the unloading facilities required for each alternative. It was hypothesized that high volume receivers can benefit from economies of scale in operating a rail siding. By the same token, low volume receivers should try to minimize the cost of owning and operating unloading facilities by using a door-to-door service like LTL. If these hypotheses are true, then the use rate variable should have a positive coefficient in the rail cost function and a negative coefficient in the LTL cost function.

Another alternative specific variable which has been used in this specification is distance. This variable was added to the specification to help correct for errors in the estimation of the level of service attributes. In many of the specifications which

were tested, the coefficient of the safety stock - stockout cost term had the wrong sign. This term increases when travel time unreliability increases, and unreliability usually increases with the length of the haul. Thus, it was hypothesized that the coefficient of the safety stock - stockout cost term was compensating for an over-estimate of the rail rate on long distance hauls. It was hoped that this problem could be mitigated by adding distance as an alternative specific variable.

The estimation results from this specification are shown in Table 6. Note that all coefficients have the expected sign except for the in-transit capital carrying cost. The high t statistic of this coefficient indicates that there is either a major error in the specification, or a major error in the measurement of the variables, or both.

Numerous other problems are also indicated by the results displayed in Table 6. Although the transport rate and the safety stock-stockout cost variables have coefficients with the proper sign, neither coefficient is significantly different than zero. Furthermore, three of the four alternative specific variables have insignificant coefficients. It must be concluded that the basic problems have not been solved by this specification of the model.

The specification shown in Table 5 was re-formulated in an effort to determine the exact problem with the level of service variables. The new specification is shown in Table 7. In this

Table 6

Estimation Results from the Logistics Cost  
Based Specification

<u>Variable</u>		<u>Coefficient</u>	<u>t_Statistic</u>
1) In-transit carrying cost	ITCC	+ .253	+4.54
2) Loss and damage	LD	-4253.2	-3.09
3) Transport rate	RATE	- .175	-0.29
4) Safety stock carrying cost and stockout cost (rail)	SSSC	- .058	-1.09
5) Constant (rail)	RC	-2.54	-7.72
6) Constant (LTL)	LC	+2.69	+4.23
7) Use rate (rail)	USE	+ .000011	+1.03
8) Use rate (LTL)	USE	- .000018	-0.77
9) Distance (rail)	RDST	+ .0020	+2.94
10) Distance (LTL)	LDST	- .00034	-0.29

no. of observations = 1430

\*

L (0) = -991. (log likelihood for coefficients of zero)

\*

L (B) = -315. (log likelihood for estimated coefficients)

<sup>2</sup>

pseudo r = .68 (explained log likelihood/total log likelihood)

specification, the transport tariff and the in-transit carrying cost are allowed to have a different coefficient in the cost function of each alternative. If the model is properly specified, and if the tariff and carrying cost are accurately measured in all cases, then this reformulation should have no significant effect on the estimates of the coefficients. The alternative specific coefficients of tariff and carrying cost should have (approximately) the same estimated values as their generic counterparts in the previous specification. However the estimation results shown in Table 8 do not bear out these expectations. Unlike the coefficient of carrying cost shown in Table 6, the three coefficients of carrying cost shown in Table 8 all have the proper sign. Furthermore, the carrying cost coefficient in the FTL cost function is significantly different than the carrying cost coefficient in either the LTL or rail cost functions.

The coefficient of the transport tariff has also been affected by the re-specification of the model. The coefficients of RATE in the FTL and LTL cost functions shown in Table 8 are more than three orders of magnitude larger than the generic coefficient of RATE shown in Table 6. Moreover, the coefficient of RATE in the rail cost function is unlike the other coefficients of RATE shown in Table 8, or the generic coefficient of RATE shown in Table 6.

In addition to the problems discussed above, the coefficient



Table 7

Re-specification of the Logistics Cost  
Based Mode Choice Model

<u>Coefficient</u>	<u>Alternative</u>		
	<u>LTL</u>	<u>FTL</u>	<u>Rail</u>
B <sub>1</sub>	0	0	ITCC
B <sub>2</sub>	0	ITCC	0
B <sub>3</sub>	ITCC	0	0
B <sub>4</sub>	LD	LD	LD
B <sub>5</sub>	0	0	RATE
B <sub>6</sub>	0	RATE	0
B <sub>7</sub>	RATE	0	0
B <sub>8</sub>	0	0	SSSC
B <sub>9</sub>	0	0	RC
B <sub>10</sub>	LC	0	0
B <sub>11</sub>	0	0	Q
B <sub>12</sub>	Q	0	0
B <sub>13</sub>	0	0	RDST
B <sub>14</sub>	LDST	0	0

Table 8

Estimation Results from the Revised  
Logistics Cost Based Model

<u>Variable</u>		<u>Coefficient</u>	<u>t_Statistic</u>
1) In-transit carrying cost (rail)	ITCC	-.233	-1.38
2) In-transit carrying cost (FTL)	ITCC	-1.066	-2.93
3) In-transit carrying cost (LTL)	ITCC	-.237	-0.37
4) Loss and damage	LD	-6337.	-2.97
5) Transport rate (rail)	RATE	-12.52	-0.88
6) Transport rate (FTL)	RATE	-200.9	-2.92
7) Transport rate (LTL)	RATE	-365.8	-0.99
8) Safety stock carrying cost and stockout cost (rail)	SSSC	+.867	+1.73
9) Constant (rail)	RC	-6.526	-7.39
10) Constant (LTL)	LC	-3.823	-0.39
11) Shipment size (rail)	Q	+.000070	+6.66
12) Shipment size (LTL)	Q	-.00059	-0.74
13) Distance (rail)	RDST	-.00032	-0.27
14) Distance (LTL)	LDST	+.0617	+4.12

no. of observations = 1430

\*  
L (0) = -991. (log likelihood for coefficients of zero)

\*  
L (B) = -140. (log likelihood for estimated coefficients)

$\frac{2}{2}$   
pseudo r = .86 (explained log likelihood/total log likelihood)

of the safety stock - stockout cost has the wrong sign in the latest set of results. Thus, it is hard to claim that the model has been improved by removing the constraints imposed by the generic use of the carrying cost and tariff variables.

#### Summary of Results from the Estimation of the Conditional Mode Choice Model

The problems with the data have made it impossible to produce a satisfactory mode choice model. It was expected that the estimated usage rates, and the loss and damage data would cause problems because they are poor descriptors of the actual situation at the disaggregate level. However the major problems appear to be caused by the rate and travel time data. The lack of information on the unreliability of truck shipments certainly aggravated the situation, as did inaccuracies in the rate models.

#### A Model of the Joint Choice of a Mode and a Shipment Size

Preliminary experiments were conducted with a multinomial logit model of the joint choice of a mode and a shipment size. The logit model was used because it is the only form of disaggregate model which has been applied extensively to multiple choice situations.

The set of alternatives used in the joint choice model includes two modes: rail and truck. Since the logit model

Table 9

Shipment Alternatives Used in the  
Mode and Shipment Size Choice Model

<u>Weight Group</u>	<u>Mean Shipment Size</u>	<u>Alternatives</u>	
		<u>Truck</u>	<u>Rail</u>
0 - 5000	2000	1	1
5000 - 10000	7500	2	2
10000 - 20000	15000	3	3
20000 - 30000	25000	4	4
30000 - 40000	35000	5	5
40000 - 50000	45000	6	6
50000 - 60000	55000	7	7
60000 - 80000	70000		8
80000 - 90000	85000		9
90000+	110000		10

Table 10

Definitions of Variables Used in the  
Specification of the Joint Choice Model

<u>Variable</u>	<u>Definition</u>
ITCC	in-transit capital carrying cost (dollar-days per pound)
LD	loss and damage (dollars per pound)
RATE	transport rate (dollars per pound)
SSSC	safety stock carrying cost and stockout cost (dollars per pound)
RC	(1 for rail, 0 otherwise)
LC	(1 for LTL, 0 otherwise)
NSSC	non-safety stock carrying cost incurred after receiving the shipment (dollars per pound)
HNDL	handling cost (frequency of shipments) (shipments per year)
VAL	commodity value (dollars if size < 5 tons, 0 otherwise)

requires the use of discrete alternatives, the range of shipment sizes has been broken into ten groups. Each group has been represented in the choice set by the average shipment size in the group. Combining the modes and average sizes results in the seventeen alternatives shown in Table 9. As explained in Chapter 5, the logit model requires that the random error in measuring the cost of any one alternative be independent of the random errors associated with all other alternatives. This assumption could be violated by using too many shipment size alternatives. Unfortunately, there is no a priori way of knowing whether a given choice set violates the independence assumption.

The set of variables used in specifying the joint choice model are listed in Table 10. They are basically the same variables used in the mode choice model, plus two additional logistics cost variables, and one new alternative specific variable. The new logistics cost variables represent the ordering and handling cost, and the capital carrying cost incurred after receiving the shipment. The ordering and handling cost has been computed in the manner described in Chapter 6. The capital carrying cost has also been computed according to the formula given in Chapter 6, except that the mean use rate of the average size purchaser has been substituted for the use rate of the actual purchaser.

The value of the commodity being shipped has been used as an alternative specific variable in the specification of the joint

choice model. This variable appears in the cost functions of alternatives with a shipment size of five tons or less. It was hypothesized that the loss and damage on high value shipments is minimized by the use of a small shipment size. However the loss and damage data fails to capture this effect because it has been averaged over all shipment sizes. Therefore, value has been added to the cost functions of the alternatives with small shipment sizes in order to compensate for the poor data. If this variable functions in the expected manner, then its coefficient should be positive.

The results from the conditional mode choice model indicate that shipment size should be tested as an alternative specific variable in the joint choice model. Unfortunately, the model estimation computer program would not converge when this variable was added to the specification. The source of the problem is partial collinearity between shipment size and other cost variables which include shipment size. Rather than deleting a logistics cost term from the model, the alternative specific shipment size variable was deleted.

When the mode and shipment size decisions are modeled jointly, it is not possible to use the tariff models to differentiate between LTL and FTL shipments. Therefore, it has simply been assumed that all truck shipments of less than ten tons experience the LTL level of service, while all shipments larger than ten tons experience the FTL level of service. This

Table 11

## Estimation Results form the Joint Choice Model

<u>Variable</u>		<u>Coefficient</u>	<u>t_Statistic</u>
1) In-transit carrying cost	ITCC	+.332	+ 7.28
2) Loss and damage	LD	-1732.	- 1.71
3) Transport rate	RATE	+34.78	+11.74
4) Safety stock carrying cost and stockout cost (rail)	SSSC	-2.101	-11.99
5) Constant (rail)	RC	-2.243	-16.32
6) Constant (LTL)	LC	+.572	+ 6.95
7) Non-safety stock carrying cost	NSSC	+.00042	+ 5.17
8) Handling cost	HN DL	+.147	+12.93
9) Value (if size < 5 tons)	VAL	+.274	+ 5.94

no. of observations = 1430

\*

L (0) = -3837. (log likelihood for coefficients of zero)

\*

L (B) = -2412. (log likelihood for estimated coefficients)

<sup>2</sup>

pseudo r = .37 (explained log likelihood/total log likelihood)



approach is probably accurate enough for preliminary experiments with the demand model. However, more sophisticated methods should be developed for the estimation of the level of service given to shipments of different sizes.

All of the logistics cost variables listed in Table 10 have been used generically in the specification of the joint model, except the safety stock carrying cost - stockout cost term. The results from estimating this specification are shown in Table 11. Note that the coefficient of RATE has the wrong sign. Similarly, the handling cost and the two capital carrying cost variables have coefficients with the wrong sign. It is difficult to attribute these problems to any one cause. However it is likely that errors in the estimated transport rates and travel times are largely to blame.

The results from the estimation of the joint choice model are very preliminary and inconclusive. At this point it is difficult to pinpoint defects in the model specification. It is also difficult to determine whether the discrete representation of shipment size has had any deleterious effects.

### Summary

Many of the estimation problems discussed in this chapter have been blamed on poor data. It appears that the methods used to estimate rates and travel times are particular sources of trouble. However, this does not imply that rates and travel

times can never be estimated accurately enough for use in demand modeling studies. It is possible that continued work in supply modeling could overcome these problems.

The most direct remedy for the data problems is to collect a new disaggregate data set for use in this type of research. However, if aggregate data must be used, then there are several steps which should be taken. First, a large data set should be used because aggregate data is less efficient for model estimation than disaggregate data. Secondly, an effort should be made to synthesize data on the distribution of those attributes which cannot be tied directly to individual shipments. The quality of the results might be improved significantly by incorporating the information from these distributions into the model.

## Chapter 9

Conclusions and Recommendations  
for Further Research

The need for a policy sensitive model of the demand for freight transportation has risen dramatically over the past few years. Planners are currently grappling with a multitude of complex problems ranging from energy conservation, to deregulation, to railroad bankruptcy. There are no clear cut solutions to these problems. It is likely that any policy action will result in both costs and benefits; both expected and unexpected effects. Unfortunately, the analytical tools which are needed for policy analysis are not available in the freight area. The existing freight demand models are insensitive to many of the policies of interest. They employ a very narrow range of variables and they predict a limited range of information. As a result, freight demand models are often shunned by policy-makers in both the public and private sectors.

One of the reasons for the current state of affairs is the lack of a sound theoretical framework. This research has demonstrated the usefulness of logistics management theory in this regard. Logistics cost calculations provide managers with a methodology for making shipment decisions. Therefore logistics

cost theory can provide modelers with considerable insights into the decision making process. Of primary importance is the fact that logistics theory makes it clear that freight decisions should not be modeled separately. Choices such as mode and shipment size are so closely related that they cannot be separated by decision-makers, and they should not be separated in the modeling process. This is a key point which has been overlooked in all of the model development work done to date.

The logistics management approach to model specification would only be marginally useful if it was not coupled with the use of a disaggregate model. Modeling the process at the micro level makes it possible to precisely measure the variables which influence each decision. As a result, a disaggregate model has the following features:

- Use of a wide range of policy sensitive variables. Many variables perform poorly in aggregate models because the aggregation process reduces the information in the data. These same variables can often be used very effectively in a disaggregate model.
- Efficient use of data. When each decision making situation is described precisely, fewer decisions need be studied in order to estimate the model parameters reliably.
- Transferability. If the model can capture all of the influences which bear on the shipment decisions, then the model can be used in any geographic locale.

In addition, the logistics cost specification of a disaggregate model results in a clear, behavioral interpretation of each

coefficient. Thus, the combination of logistics theory and disaggregate modeling can lead to an analysis tool which is far superior to those presently available.

It should be noted that the usefulness of the type of model proposed in this thesis is enhanced greatly by the use of a broad range of level of service attributes and commodity attributes. Most government and carrier policies are designed to influence the level of service. Hence the more level of service variables in the model, the greater the variety of policies which can be analyzed. However, the model would be of little value if it could be applied only to the commodities represented in the estimation data set. The use of commodity attributes in the specification makes it possible to apply the estimated model to the full range of commodities.

This research has been aimed at the development of a flexible freight demand model for use in policy analysis. However, there are two important aspects of policy analysis which have not been addressed. First, there is the problem of translating policies into changes in the level of service. How will deregulation of the carriers affect rates? And how will the elimination of classification yards affect rail reliability? A methodology for answering these kinds of questions must be developed before the full potential of freight demand models can be realized. It is also necessary to find better techniques for making aggregate predictions with disaggregate models. Koppelman

(1975) has explored the strengths and weaknesses of several methods of forecasting with disaggregate models. His results indicate that reliable methods require data sets which are not significantly larger than those needed to apply aggregate models. Clearly further research is needed into the impact of policies on the level of service, and into the problem of forecasting. However, these subjects fall outside the scope of this thesis. Thus, the discussion will now turn to recommendations for further research in freight demand modeling.

#### Directions for Further Research

There are basically two obstacles to the creation of the type of demand model discussed in this thesis. One is the collection of the proper disaggregate data and the other is the development of a disaggregate model that can handle the joint choice of discrete and continuous variables.

At the present time, the data problem is the most imposing constraint on model development. The data which was used in this thesis was insufficient for the estimation of a useful disaggregate demand model. Surprisingly, the principal deficiencies are not in the flow data, but in the level of service attributes and receiver attributes. It appears that the most efficient approach to resolving these problems is to collect a new disaggregate data base. This data base should include an extensive variety of level of service attributes, commodity

attributes, receiver attributes, and market attributes. Most importantly, the data should fully describe the supplier, mode and shipment size alternatives available to each decision-maker. It is not essential that the sample cover a huge number of decision-makers, although each observation should be covered in detail.

The other problem which needs to be tackled is the development of a new form of disaggregate model. This model must meet three criteria. First, it must be capable of handling the joint choice of two or more discrete variables and at least one continuous variable. Secondly, the choice set must be able to accommodate three or more alternatives for each of the discrete choice variables. Thirdly, the model must be compatible with the principle of cost minimization. As pointed out in Chapter 5, none of the existing models meet all three criteria. Nevertheless, recent work by Westin (1975) has produced some interesting results which may eventually lead to the type of model which is desired.

In summary, this research has resulted in a sound theoretical framework which should be helpful in future model estimation work. Unfortunately, the model which has been proposed could not be implemented with the resources and data available for this thesis. However, the approach that has been developed is feasible. Further research into improved data sets and specifications will produce the tools needed for freight policy analysis within the next three to five years.



Bibliography

1. A.D. Little Inc., Domestic Waterborne Shipping Market Analysis, Appendix A - Development of the Forecasting Data Base; Appendix B - Forecasting Methodology; Appendix C - Modal Share Estimates, 1974.
2. Allen, W., "The Impact of Loss and Damage on Transport Demand", Proceedings of the Transportation Research Forum, 1973.
3. American Airlines, "Distribution Logistics Marketing", 1975.
4. American Trucking Association, American Trucking Trends, Washington, D.C., 1967.
5. Antle, L. and Haynes, R., "An Application of Discriminant Analysis to the Division of Traffic Between Transport Modes", Report 71-2, U.S. Army Engineer Institute for Water Resources, 1971.
6. Army Corps of Engineers, Waterborne Commerce Statistics of the United States, U.S. Government Printing Office, Washington, D.C., 1972.
7. Association of American Railroads, "Final Report on a Pilot Study of Rail Freight Car Movements and Shipper's Criteria for Selecting Transport Modes", A.A.R., Washington, D.C., 1971.
8. Bayliss, B., "Demand for Freight Transportation - Practical Results of Studies on Market Operations", Report on the Twentieth Round Table on Transport Economics, European Conference of Ministers of Transport, Paris, 1973.
9. Beuthe, M. and Moses, L., "The Demand for Freight Transportation: The Influence of Time", in Transportation: A Service, Academy of Science, New York, 1968.
10. Beuthe, M., "Freight Transportation Mode Choice: An Application to Corn Transportation", Ph.D. dissertation, Northwestern University, 1968.
11. Beuthe, M., "A Predictive Model of Regional Demands for Freight Transportation", Journal of Regional Science, Volume 12, no. 1, 1972.

12. Black, W., "Interregional Commodity Flows: Some Experiments with Gravity Models", *Journal of Regional Science*, April, 1972.
13. Boeing Commercial Airplane Company, "United States Domestic Air Freight: An Econometric Forecast", Seattle, Washington, 1974.
14. Buffa, E. and Taubert, W., Production - Inventory Systems: Planning and Control, Richard Irwin, 1972.
15. Buhl, W., "Intercity Highway Transport Share Tends to Vary Inversely with Size of Plant", *Highway Research Record*, no. 175, Transportation Research Board, 1967.
16. Charles River Associates, Competition Between Rail and Truck in Intercity Freight Transportation, NTIS PB 193 354, 1969.
17. Chung, C. and Roberts, P., "Design of a Structure and Data Analysis Scheme for Intercity Freight Demand Forecasting", Report 75-15, Center for Transportation Studies, Massachusetts Institute of Technology, 1975.
18. Church, D., "Impacts of Size and Distance on the Intercity Highway Share of Transportation of Industrial Products", *Highway Research Record*, no. 175, Transportation Research Board, 1967.
19. Department of Commerce (Bureau of the Census), Census of Agriculture, Volumes 1 through 6, U.S. Government Printing Office, 1972.
20. Department of Commerce (Bureau of the Census), 1967 Census of Manufacturers, U.S. Government Printing Office, Washington, D.C., 1971.
21. Department of Commerce (Bureau of the Census), 1967 Census of Transportation, U.S. Government Printing Office, Washington, D.C., 1971.
22. Department of Commerce (Bureau of the Census), County and City Data Book, 1972, U.S. Government Printing Office, 1973.
23. Department of Commerce (Bureau of the Census), County Business Patterns, 1967, U.S. Government Printing Office, 1968.

24. Department of Transportation, Carload Waybill Statistics, 1969, Office of Systems Analysis and Information, Department of Transportation, Washington, D.C., 1971.
25. Gaudio S. and Meko E., "A Study of Plant Characteristics and Factors Which Lend Themselves to an Explanation of Goods Movements at Industrial Sites", S.M. thesis, The Transportation Center, Northwestern University, 1973.
26. Graves, F., "The Shipper Preference Model", The Proceedings of the Transportation Research Forum, 1966.
27. Haefner, L., "Analytic Issues in Goods Movements", presented at the ORSA/TIMS Joint National Meeting, Boston, 1974.
28. Hartwig, J. and Linton, W., "Disaggregate Mode Choice Models of Intercity Freight Movement", S.M. thesis, The Transportation Center, Northwestern University, 1974.
29. Heredeen, J., "Theoretical Development and Preliminary Testing of a Mathematical Model for Predicting Modal Split", Report TTSC 6908, Department of Civil Engineering, Penn. State University, 1969.
30. Institute for Water Resources, "Discriminant Analysis Applied to Commodity Shipments in the Arkansas River Area", U.S. Army Corps of Engineers, Southwest Division, 1974.
31. Interstate Commerce Commission (Bureau of Accounts), Freight Commodity Statistics, Class I Railroads, Calendar Year 1973, U.S. Government Printing Office, Washington, D.C., 1974.
32. Interstate Commerce Commission (Bureau of Accounts), Freight Commodity Statistics of Motor Carriers of Property, Calendar Year 1973, U.S. Government Printing Office, Washington, D.C., 1974. Nqton, D.C., 1974.
33. Interstate Commerce Commission, Transport Statistics in the United States, 1971, Parts 1 through 9, U.S. Government Printing Office, Washington, D.C., 1973.
34. Jack Faucett Associates, Projections of Rail Freight Demand in the Conrail Region, U.S. Railway Association, 1974.
35. Komatsu, B., "Factors Influencing Freight Transportation Mode Choice", Proceedings of the International Conference on Transportation Research, Bruges, Belgium, 1973.
36. Kresge, D. and Roberts, P., Systems Analysis and Simulation, Volume 2 in the series: Techniques of Transport Planning, The Brookings Institute, 1971.

37. Kullman, B. "A Model of Rail/Truck Competition in the Intercity Freight Market", Studies in Railroad Operations and Economics, Volume 15, Massachusetts Institute of Technology, 1973.
38. Martland, C., "Procedures for Improving Railroad Reliability", Studies in Railroad Operations and Economics, Volume 12, Department of Civil Engineering, Massachusetts Institute of Technology, 1974.
39. Mathematica, Studies on the Demand for Freight Transportation, Volumes I, II, III, 1967, 1969, 1969 (respect.).
40. Mera, K., "An Evaluation of Gravity and Linear Programming Models for Predicting Interregional Commodity Flows" in Volume 1 of Techniques of Transport Planning, The Brookings Institute, 1971.
41. Meyer, J.; Peck, M.; Stenason, J. and Zwick, C., The Economics of Competition in the Transportation Industries, Harvard University Press, 1959.
42. Miklius, W., "Estimating the Freight Traffic of Competing Modes: An Application of the Linear Discriminant Function", Land Economics, May, 1969.
43. Miller, E., "Effect of Rates on the Demand for Railroad Services", Proceedings of the Transportation Research Forum, 1972.
44. Morton, A., "A Statistical Sketch of Intercity Freight Demand", Highway Research Record, no. 296, Transportation Research Board, 1969.
45. Morton, A., "Competition in the Intercity Freight Market: A Waybill Study of the Motor Carrier Industry", Office of Systems Analysis, Department of Transportation, Washington, D.C., 1971.
46. Morton, A., "Freight Demand", Ph.D. dissertation, Economics Department, Harvard University, 1973.
47. Morton, A., "Intermodal Competition for the Intercity Transport of Manufactures", Land Economics, November, 1972.

48. Morton, A., "Truck-Rail Competition for Traffic in Manufactures", Proceedings of the Transportation Research Forum, 1971.
49. Moses, L. et. al., Cost Benefit Analysis for Inland Navigation Improvements, Report 70-4, U.S. Army Engineer Institute for Water Resources, 1970.
50. Niedercorn, J. and Moorehead, J., "The Commodity Flow Gravity Model - A Theoretical Reassessment", Regional and Urban Economics, no. 4, 1974.
51. Perle, E., The Demand for Transportation: Regional and Commodity Studies in the United States, University of Chicago Press, 1964.
52. Reebie Associates, "Truck/Rail Intermodal Operations: An Option for the Future?", Department of Transportation, Federal Highway Administration, 1974.
53. Roberts, P., "Factors Influencing the Demand for Goods Movement", presented at the Annual Meeting of the Transportation Research Board, Washington, D.C., 1976.
54. Roberts, P., "The Logistics Management Process as a Model of Freight Traffic Demand", Proceedings of the International Symposium on Freight Traffic Models, Amsterdam, 1971.
55. Roberts, P. and Ginn, J., "Stockout Costs in Inventory Management", Harvard Business School Working Paper 71-9, Harvard University, 1971.
56. Ruijgrok, C. and van Es, J., "Modal Split in Freight Transportation", Proceedings of the International Conference on Transportation Research, Bruges, Belgium, 1973.
57. Sasaki, B., "Predicting Transporter's Choice of Mode", presented to the Annual Meeting of the Transportation Research Board, 1974.
58. Sloss, J., "The Demand for Intercity Motor Freight Transport: A Macroeconomic Analysis", The Journal of Business, University of Chicago, January, 1971.
59. Stucker, J., An Econometric Model of the Demand for Transportation, Ph.D thesis, Northwestern University, 1970.

60. Surti, V. and Ebrahimi, A., "Modal Split of Freight Traffic", Traffic Quarterly, October, 1972.
61. Swerdloff, C., "Developing a National Network Model of Intercity Freight Movement in the United States", Proceedings of the International Symposium on Freight Traffic Models, Amsterdam, 1971.
62. Temple, Barker and Sloane Inc., Forecasts of Traffic and Revenues 1974-1980, U.S. Railway Association, 1974.
63. Terziev, M.; Ben-Akiva, M. and Roberts, P., "Freight Demand Modeling: A Policy Sensitive Approach", Center for Transportation Studies Report 75-6, Massachusetts Institute of Technology, 1975.
64. Tihansky, D., "Methods of Estimating the Volume and Energy Demand of Freight Transportation", Report R-988-NSF, Rand Corp., 1972.
65. Trinc Transportation Consultants, Trinc's Blue Book of the Trucking Industry, Washington, D.C., 1974
66. Trinc Transportation Consultants, Trinc's Red Book of the Trucking Industry, Washington, D.C., 1974
67. Wallace, R., A Domestic Multi-Modal Goods Distribution Model with Emphasis on Air Cargo, NTIS N72 21969, 1971.
68. Wang, G. and Epstein, R., "Econometric Models of Aggregate Freight Transport Demand", Transportation Systems Center working paper WP-210-U1-81-A, Department of Transportation, Cambridge, Mass., 1975.
69. Watson, P.; Hartwig, J. and Linton, W., "Factors Influencing Shipping Mode Choice for Intercity Freight: A Disaggregate Approach", Proceedings of the Transportation Research Forum, 1974.
70. Westin, R., "An Econometric Model of Inter-Related Discrete and Continuous Choices", presented at the Third World Congress of the Econometric Society, Toronto, 1975.
71. Whitten, H., "Transport Flow Data: Proceedings of the National Transportation Flow Statistics Forum", The Transportation and Logistics Research Center, School of Business Administration, The American University, 1968.

## Appendix I

## Other Sources of Data

The scarcity of data is responsible in large part for the absence of many important variables from the freight demand models which have been published to date. However, several groups of researchers have partially overcome this constraint by collecting data for use in their freight demand studies.

Antle and Haynes (1971) collected a small disaggregate data set (87 observations) on barge, truck, rail and combined barge-rail movements in the Ohio River Valley. They concentrated their survey on firms receiving shipments of coal, coke, chemicals and petroleum. From each receiver interviewed, they collected data on the shipment of one commodity by one mode, from one shipper. The characteristics included in the data are:

- annual tonnage of the commodity shipped
- distance
- average travel time
- average shipment size
- rate
- handling cost
- rate on the unchosen mode

Unfortunately, they did not collect data on the average travel time, shipment size, and handling cost on the unchosen mode.

The Army Corps of Engineers-Southwest Division has conducted a study similar to the Antle and Haynes study. They collected disaggregate data on 195 barge, rail and truck shipments of a variety of goods in the Arkansas River Valley. This data set includes the same variables as the Antle and Haynes data except that the rate on the unchosen mode was not recorded.

Brian Kullman (1973) built a data set around the commodity flow information contained on the Census of Transportation computer tapes. He used the Carload Waybill Statistics to develop regression equations that can be used to estimate rail rates for certain commodities as a function of distance. He also used truck rate data from a paper by Morton (1971) to calibrate regression equations that can be used to estimate truck rates as a function of distance. Kullman obtained average rail transit times for city pairs in the Northeast from Penn Central records. He used the same records to calculate the travel time time reliability for city pairs. He was unable to obtain similar information for trucks, and therefore he estimated the truck transit times based on an assumed average daily mileage. The truck travel time reliability was assumed to be unity (i.e.,



perfect reliability). Kullman used Census of Manufacturers data to estimate the value of commodities. He simply divided the value of total output by the volume of output for selected commodity groups.

Kullman's attempts to use this data base to estimate aggregate mode choice models, as described in Chapter 4, met with little success. He attributed many of the problems to measurement errors and the aggregate nature of this data set.

Hartwig and Linton (1974) collected 1213 freight waybills for full load shipments of a particular consumer durable which were sent by rail and truck. From the bills, they determined the distance shipped, travel time, cost, shipment size and the value of the commodity being shipped. The relatively good empirical results reported by these researchers demonstrates the usefulness of accurate disaggregate data.

The Chicago Area Transportation Study has conducted a survey of firms which only have truck service available. This data set contains very little level of service information. However, the description of the shipment and the shipper/receiver attributes that were collected are uncommonly detailed.

The list of shipper/receiver attributes that were collected includes the following:

- 1) Floorspace Variables
  - a. Office floorspace
  - b. Manufacturing floorspace
  - c. Total storage space
  - d. Total plant floorspace
  
- 2) Employment Variables
  - a. Managerial employees
  - b. Sales personnel
  - c. Skilled employees-Manufacturing
  - d. Unskilled employees-Manufacturing
  - e. Service employees
  - f. Other employees
  - g. Total employees
  
- 3) Dummy Variables
  - a. Adequacy of small letter storage space
  - b. Private truck availability
  - c. Seasonal fluctuation of outbound shipments
  - d. Seasonal fluctuation of inbound shipments

The data collected on individual shipments include the following:

- 1) Origin and Destination
  
- 2) For each commodity in the shipment:
  - a. Volume
  - b. Weight
  - c. Packaging
  - d. Handling
  - e. Value
  
- 3) For the total shipment:
  - a. Volume
  - b. Weight
  - c. Value
  - d. Transportation cost

The list of shipper/receiver attributes and shipment attributes covered in this survey could be expanded somewhat. But even so, this survey is much more detailed and comprehensive than most.

None of the existing disaggregate data sets contain enough information to allow the estimation of a complete disaggregate freight demand model. Most are based on a small number of commodities. Also, most of these data sets contain very few receiver attributes, market attributes or transport level of service attributes. The quality of these data make it difficult to fully test the type of specification discussed in Chapter 6.

## Appendix II

## Production Areas and Market Areas

The following is a list of the Standard Metropolitan Statistical Areas included in each of the Production Areas and Market Areas used in the 1967 Census of Transportation O-D File 1 computer tape.

1. Boston, Worcester, Providence-Pawtucket-Warwick, Brockton, Lawrence-Haverhill, Lowell
2. Hartford, New Britain, Meriden, Waterbury, New Haven, Bridgeport, Norwalk, Stamford, Springfield-Chicopee-Holyoke
3. New York, New York
4. Newark, Jersey City, Paterson-Clifton-Passaic, Middlesex County, Somerset County
5. Philadelphia, Wilmington, Trenton
6. Baltimore
7. Allentown-Behtlehem-Easton, Reading
8. Harrisburg, Lancaster, York
9. Syracuse, Utica-Rome, Albany-Schenectady-Troy
10. Buffalo, Rochester
11. Cleveland, Akron, Canton, Loraine-Elyria, Youngstown-Warren, Erie
12. Pittsburg, Steubenville-Weirton, Wheeling
13. Detroit, Flint, Toledo, Ann Arbor
14. Cincinnati, Dayton, Hamilton-Middletown, Springfield
15. Chicago, Gary-Hammond-East Chicago
16. Milwaukee, Kenosha, Racine
17. Minneaposis-St. Paul
18. St. Louis
19. Atlanta
20. Dallas, Fort Worth
21. Houston, Beaumont-Port Arthur, Galveston-Texas City
22. Denver
23. Seattle-Everett, Tacoma
24. San Francisco-Oakland, Vallejo-Napa, San Jose

25. Los Angeles-Long Beach, Anaheim-Santa Ana-Garden Grove, San Bernardino-Riverside-Ontario
31. Scranton, Wilkes-Barre-Hazelton, Pa.,  
Binghamton, N.Y.
32. Washington, D.C., Md., Va.
33. Newport News-Hampton, Norfolk-Portsmouth, Va.
34. Columbus, Ohio
35. Grand Rapids, Muskegon-Muskegon Heights, Mich.
36. Indianapolis, Muncie, Terre Haute, Indiana
37. Louisville, Kentucky,-Indiana
38. Nashville, Tenn.
39. Memphis, Tenn.-Ark.
40. Augusta, Ga.-S.C., Columbia, S.C.
41. Ft. Lauderdale-Hollywood, Miami, West Palm Beach, Fla.
42. Birmingham, Tuscaloosa, Ala.
43. Tampa-St. Petersburg, Fla.
44. Mobile, Ala., Pensacola, Fla.
45. New Orleans, Louisiana
46. Omaha, Lincoln, Nebraska
47. Kansas City, Mo.-Kans., St. Joseph, Mo., Topeka, Kans.
48. Oklahoma City, Tulsa, Oklahoma
49. San Antonio, Austin, Texas
50. Salt Lake, Provo-Orem, Ogden, Utah
51. Phoenix, Tucson, Ariz.
52. Portland, Oregon
53. Sacramento, Stockton, Calif.
54. Fresno, Bakersfield, Calif.
55. San Diego, Calif.

It should be noted that regions 1 through 55 are Market Areas, while only regions 1 through 25 are Production Areas.

## Appendix III

## Shipment Sizes

The following is a list of the shipment size groupings used in the 1967 Census of Transportation O-D File 1 computer tape. Next to each grouping is the mean shipment size which has been assumed for the purposes of the empirical work described in Chapter 8.

<u>Weight block (lbs)</u>	<u>Assumed mean</u>
under 50	30
50 - 99	75
100 - 199	150
200 - 499	350
500 - 999	750
1,000 - 2,999	2,000
3,000 - 4,999	4,000
5,000 - 9,999	7,500
10,000 - 19,999	15,000
20,000 - 29,999	25,000
30,000 - 39,999	35,000
40,000 - 49,999	45,000
50,000 - 59,999	55,000
60,000 - 79,999	70,000
80,000 - 89,999	85,000
90,000 - 99,999	95,000
100,000 - 119,999	110,000
120,000 - 149,999	135,000
150,000 - 199,999	175,000
over 200,000	220,000