Research Report No. UVACTS-14-0-85
August, 2005

Supply Chain Models for Freight Transportation Planning

University of Virginia

By:

Vidya Charan Tatineni

Dr. Michael J. Demetsky

A Research Project Report
For the Mid-Atlantic Universities Transportation Center (MAUTC)
A U.S. DOT University Transportation Center

Dr. Michael J. Demetsky
Department of Civil Engineering
Email: mjd@virginia.edu

Center for Transportation Studies at the University of Virginia produces outstanding transportation professionals, innovative research results and provides important public service. The Center for Transportation Studies is committed to academic excellence, multi-disciplinary research and to developing state-of-the-art facilities. Through a partnership with the Virginia Department of Transportation's (VDOT) Research Council (VTRC), CTS faculty hold joint appointments, VTRC research scientists teach specialized courses, and graduate student work is supported through a Graduate Research Assistantship Program. CTS receives substantial financial support from two federal University Transportation Center Grants: the Mid-Atlantic Universities Transportation Center (MAUTC), and through the National ITS Implementation Research Center (ITS Center). Other related research activities of the faculty include funding through FHWA, NSF, US Department of Transportation, VDOT, other governmental agencies and private companies.

CTS Website	Center for Transportation Studies
http://cts.virginia.edu	University of Virginia
	351 McCormick Road, P.O. Box 400742
	Charlottesville, VA 22904-4742
	434.924.6362

1. Report No. UVACTS-14-0-85	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Supply Chain Models for Freight Transportation Planning		5. Report Date August, 2005 6. Performing Organization Code
7. Author(s) Vidya Charan Tatineni Dr. Michael J. Demetsky		8. Performing Organization Report No.
9. Performing Organization and Address Center for Transportation Studies University of Virginia PO Box 400742 Charlottesville, VA 22904-7472		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agencies' Name and Address Office of University Programs, Research Innovation and Technology Administration US Department of Transportation 400 Seventh Street, SW Washington DC 20590-0001		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study investigates the applicability of a supply chain based modeling methodology for regional freight transportation planning. This methodology attempts to relate the supply chain practices of individual firms to public sector transportation planning. A two-step methodology that makes use of some of the supply chain characteristics is proposed for freight transportation planning. The first step of the methodology is to obtain the O-D Flows by tracing the supply chains of major business units in a region. This step is illustrated using the sales volume data of a truck manufacturer in Virginia. The second step is to model the choice of mode for freight shipments. The logistical needs and constraints of a shipper determine the choice of mode. Therefore, a model that accounts for the logistical variables would be appropriate for modeling the choice of mode. A list of supply chain variables that have the potential to influence the choice of mode is identified. A common problem that is usually reported in modeling the choice of mode is the lack of availability of reliable disaggregate data. An attempt has been made to develop a mode choice model using aggregate data from TRANSEARCH database supplemented with data from a survey of shippers. This survey also colleted data pertaining to relative weights among potential attributes that affect the choice of mode for three different categories of shippers. The mode choice model was developed using four different classification methods, namely: Binary Logit Model, Linear Discriminant Analysis, Quadratic Discriminant Analysis and Tree Classification. The advantages and disadvantages of using these methods for mode choice analyses are discussed.

17 Key Words Freight Transportation Planning, Supply Chain Model	18. Distribution Statement No restrictions. This document is available to the public.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 98	22. Price N/A

ACKNOWLEDGEMENTS

I would like to thank the Mid-Atlantic Universities Transportation Center for supporting this research effort. I will always be thankful to my advisor Prof. Michael Demetsky for his continuous guidance and support through out the course of my study at the University of Virginia. I greatly appreciate Prof. Lester Hoel for his wonderful course on Intermodal Transportation and for his advice on developing my presentation skills. I would like to thank John Miller and Roger Howe of the Virginia Transportation Research Council for their help in preparing the survey and for their feedback on this project. I would also like to thank the other faculty at the University of Virginia, especially Professors Harry Teng and Donald Brown for teaching me two very important courses on choice modeling techniques. Finally, I wish to thank all my friends and family members for their support and encouragement throughout my stay in the United States.

ABSTRACT

This study investigates the applicability of a supply chain based modeling methodology for regional freight transportation planning. This methodology attempts to relate the supply chain practices of individual firms to public sector transportation planning. A two-step methodology that makes use of some of the supply chain characteristics is proposed for freight transportation planning. The first step of the methodology is to obtain the O-D Flows by tracing the supply chains of major business units in a region. This step is illustrated using the sales volume data of a truck manufacturer in Virginia. The second step is to model the choice of mode for freight shipments. The logistical needs and constraints of a shipper determine the choice of mode. Therefore, a model that accounts for the logistical variables would be appropriate for modeling the choice of mode. A list of supply chain variables that have the potential to influence the choice of mode is identified. A common problem that is usually reported in modeling the choice of mode is the lack of availability of reliable disaggregate data. An attempt has been made to develop a mode choice model using aggregate data from TRANSEARCH database supplemented with data from a survey of shippers. This survey also colleted data pertaining to relative weights among potential attributes that affect the choice of mode for three different categories of shippers. The mode choice model was developed using four different classification methods, namely: Binary Logit Model, Linear Discriminant Analysis, Quadratic Discriminant Analysis and Tree Classification. The advantages and disadvantages of using these methods for mode choice analyses are discussed.

TABLE OF CONTENTS

CHAPTER 1 1

INTRODUCTION 1

1.1 Introduction 1
1.2 Logistics and Supply Chain Management 3
1.3 Changes Taking Place in Logistics Practices 3
1.3.1 Shift from "Push" to "Pull" Logistics 4
1.3.2 Traditional "Push" Logistics System 4
1.3.3 Modern "Pull" Logistics System 5
1.3.4 Emergence of Electronic Commerce 6
1.4 Demand Management Efforts 6
1.5 Problem Statement 7
1.6 Purpose and Scope 8

CHAPTER 2 9

LITERATURE REVIEW 9

2.1 Introduction 9
2.2 Problems with Existing Freight Planning Methodologies 9
2.3 Logistics Costs 10
2.4 Trends in Logistics Costs 11
2.5 Factors Affecting Total Logistics Cost to Sales Ratio 12
2.6 Service Related Costs 14
2.7 Comparison of Truck and Rail 14
2.8 Freight Transportation Planning Models 17
2.8.1 Trip Generation and Distribution Modeling 17
2.8.2 Mode Choice Modeling 18
2.9 Summary 20

CHAPTER 3 21

METHODOLOGY 21

3.1 Need for a Supply Chain Based Modeling Methodology 21
3.2 Problems with the Conventional Data Sources 22
3.3 Proposed Modeling Methodology 23
3.4 Illustration of the Methodology 24

CHAPTER 4 27

VOLVO TRUCKS CASE STUDY 27

4.1 Corporate History.. 27

4.2 The Volvo Supply Chain 27

4.2.1 Suppliers 27

4.2.2 Dealer Network.. 28

4.2.3 Supply Chain Management... 28

4.3 Volvo Sales and Market Share.. 30

4.4 Comparison of Flows with TRANSEARCH Data..................................... 30

4.5 Summary 31

CHAPTER 5 33

STUDY OF FACTORS AFFECTING THE CHOICE OF MODE.......33

5.1 Identification of Supply Chain Variables to be Studied 33

5.1.1 Shipper characteristics 33

5.1.2 Commodity characteristics.. 34

5.1.3 Logistic characteristics... 34

5.1.4 Modal characteristics 34

5.2 Design of the Questionnaire.. 35

5.3 Recipients of the Survey 36

5.4 Summary of Survey Responses 37

5.5 Relative Preferences by Commodity Type 40

5.6 Performance of Truck versus Rail 43

5.7 Summary 44

CHAPTER 6 46

EMPIRICAL CHOICE MODELING..46

6.1 Need for Empirical Choice Modeling.. 46

6.2 Training Data Set and Test Data Set.. 47

6.3 Preparation of Data Sets... 47

6.4 Development of a Binary Logit Model for Choice of Mode 50

6.5 Mode Choice Modeling using Linear Discriminant Analysis (LDA)................. 56

6.5.1 Model I... 57

6.5.2 Model II 58

6.6 Mode Choice Modeling using Quadratic Discriminant Analysis (QDA)........... 58

6.6.1 Model I... 59

6.6.2 Model II 59

6.7 Mode Choice Modeling using Tree Based Methods.................................. 60

6.7.1 Tree pruning using cross-validation........ 61
6.8 Summary........ 63

CHAPTER 7 65

CONCLUSIONS 65

7.1 Summary........ 65
7.2 Conclusions........ 68
7.3 Applications for Statewide Freight Transportation Planning 69
7.4 Recommendations for Future Research........ 70

REFERENCES 71

APPENDIX A: PRELIMINARY QUESTIONNAIRE 76

APPENDIX B: ACTUAL QUESTIONNAIRE USED FOR THE SURVEY........ 83

APPENDIX C: SAMPLE 'R' CODES USED FOR MODELING........ 87

APPENDIX D: 'R' OUTPUT FOR TREES 96

LIST OF TABLES

Table 4.1: Number of Tons of STCC 3711 Shipped from Virginia's Pulaski County to Each State in the U.S. for 2003……………………………………………………………29

Table 4.2: Annual Market Share and Sales Information for Volvo Trucks……………….30

Table 4.3: Comparison of Estimated Flows with TRANSEARCH Flows for 1998……….31

Table 5.1: Relative Weights of Attributes for All Shippers, Shippers Using Only Truck and Shippers Using Both Truck and Rail…………………………………………………....37

Table 5.2: Relative Weights of Attributes for All Shippers by Commodity Type………...37

Table 5.3: Relative Weights of Attributes by Commodity Type for Shippers Using Only Truck and Shippers Using Both Truck and Rail…………………………………………….38

Table 5.4: Comparison of Travel Time and On-Time Performance for Truck and Rail....43

Table 5.5: Commodity Wise Comparison of Transportation and Other Logistics Costs for Truck and Rail……………………………………………………………………………..44

Table 6.1: Logit Model Parameter Estimates for Model I…………………………………..51

Table 6.2: Accuracy of Logit Model I with a probability threshold of 0.50………………..52

Table 6.3: Accuracy of Logit Model I with a probability threshold of 0.75………………..52

Table 6.4: Correlation Matrix for All the Explanatory Variables…………………………53

Table 6.5: Logit Model Parameter Estimates for Model II…………………………………53

Table 6.6: Accuracy of Logit Model II with a probability threshold of 0.50………………54

Table 6.7: Accuracy of Logit Model II with a probability threshold of 0.75………………54

Table 6.8: Distances at Which Shippers Begin to Prefer Rail for Various Product Values and Annual Tonnages…………………………………………………………………………55

Table 6.9: Co-efficients of Linear Discriminants for Model I………………………………57

Table 6.10: Accuracy of LDA Model I with default prior probabilities π_k and π_l…………57

Table 6.11: Accuracy of LDA Model I with probabilities $\pi_k = 0.25$ and $\pi_l = 0.75$…………57

Table 6.12: Co-efficients of Linear Discriminants for Model II……………………………58

Table 6.13: Accuracy of LDA Model II with default prior probabilities π_k and π_l………..58

Table 6.14: Accuracy of LDA Model II with probabilities $\pi_k = 0.25$ and $\pi_l = 0.75$………...58

Table 6.15: Accuracy of QDA Model I with default prior probabilities π_k and π_l………...59

Table 6.16: Accuracy of QDA Model II with default prior probabilities π_k and π_l………..59

Table 6.17: Prediction Accuracy of a Five Node Tree………………………………………62

Table 6.18: Comparison of Prediction Accuracy of all Four Methods on Training Set…..64

Table 6.19: Comparison of Prediction Accuracy of all Four Methods on Test Set………..64

LIST OF FIGURES

Figure 1.1: Steps in the Statewide Intermodal Freight Transportation Methodology………………………………………………………………………….…….. 2

Figure 1.2: A Typical Supply Chain…………………………………………….………. 4

Figure 2.1: Components of Logistics Cost………………………………………………10

Figure 2.2: Annual Trend in Logistics Cost to Sales Ratio……………………………......12

Figure 2.3: Variation of Logistics Cost to Sales Ratio with Company Size…………………13

Figure 2.4: Variation of Logistics Cost to Sales Ratio with Product Value…………………14

Figure 2.5: Modal Comparison of On-Time Delivery Performance………………...........15

Figure 2.6: Modal Comparison of Freight Loss and Damage………………………………16

Figure 2.7 Modal Comparison of Equipment Availability…………………………….……16

Figure 5.1: Average Relative Weights of Attributes for All the Shippers…………….……39

Figure 5.2: Average Relative Weights of Attributes for Shippers Using only Truck……....39

Figure 5.3: Average Relative Weights of Attributes for Shippers Using Truck and Rail………………………………………………………………………………………40

Figure 5.4: Commodity wise Average Relative Weights of Attributes for All the Shippers…………………………………………………………………………………....41

Figure 5.5: Commodity wise Relative Weights of Attributes for Shippers using only Truck… …………………………………………………………………………….....42

Figure 5.6: Commodity wise Relative Weights of Attributes for Shippers Using Both Truck and Rail……………………………………………………………………………42

Figure 6.1: A Fully Grown Tree…………………………………………………………61

Figure 6.2: Mis-Classification Rate versus Tree Size………………………......................62

Figure 6.3: A Five Node Classification Tree………………………………………………63

Chapter 1

Introduction

1.1 Introduction:

The mobility of freight is vital to the national economy. It is estimated that about 11.6 billion tons of freight, which is worth $ 8.4 trillion, moved within and across the U.S. in 2002 [1]; and the amount of freight movement is expected to increase by about 70% by the year 2020. The growth in demand for freight transportation has already outgrown the infrastructure improvements taking place to accommodate the growth at many places. The problem is more acute on the highway system in metropolitan areas where severe congestion has reduced the efficiency of the freight transportation system.

Because of the importance of freight movement in economic development, there has been an increased attention towards incorporating freight into the transportation planning process. Both the Intermodal Surface Transportation Equity Act of 1991 (ISTEA) and the Transportation Equity Act for the 21^{st} Century (TEA-21) of 1998 require State Departments of Transportation (DOTs) and Metropolitan Planning Organizations (MPOs) to consider freight movement in their planning process.

In compliance with ISTEA and TEA-21, the Virginia Transportation Research Council (VTRC) developed a Statewide Intermodal Freight Transportation Planning Methodology for Virginia [2]. This methodology has proposed a six step planning process for Virginia. The six steps of the planning process are shown in Figure 1.1.

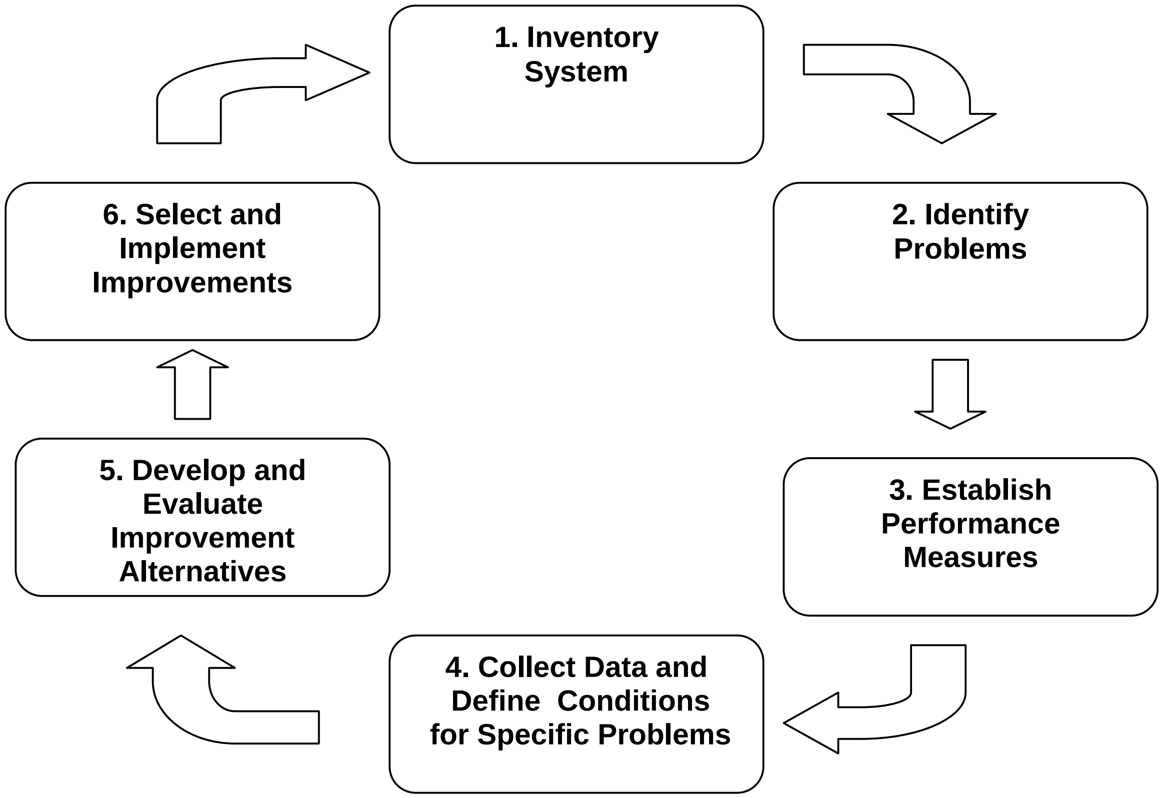


Figure 1.1: Steps in the Statewide Intermodal Freight Transportation Methodology
Source: Reference [2]

The first step of this planning process is the Inventory System step and this involves taking an inventory of the existing freight infrastructure and obtaining the freight flows by commodity and mode. This step is the most crucial and expensive step in this methodology as it serves as an input to all the other steps.

As a part of this step, a list of key commodities that are deemed important for Virginia's freight transportation were identified in a previous study [3]. As a part of this study, the trip production and attraction equations were developed to facilitate the forecasting of future freight flows. Another study was done to distribute the trips that belong to the truck mode [4]. As a continuation of the previous two studies, the present study aims at incorporating the logistical characteristics of the supply chains into the freight planning process.

1.2 Logistics and Supply Chain Management

Many times the words Logistics Management and Supply Chain Management are used interchangeably. However, there is a difference in scope between these two practices. Logistics Management can be defined as the process of managing the physical distribution of goods in a firm. This involves managing the inbound and outbound movements along with the inventory of the firm. Supply Chain Management encompasses a broader set of functions including Demand Forecasting, Sourcing and Procurement, Coordinating the Manufacturing Activities and Logistics Management. Supply Chain Management can be considered as an evolution from Logistics Management and many firms are shifting from the practice of Logistics Management to Supply Chain Management because it brings a greater amount of coordination between the various elements of the supply chain. As a consequence of this trend the Council of Logistics Management (CLM) has been renamed as Council of Supply Chain Management Professionals (CSCMP) on January 1st 2005.

1.3 Changes Taking Place in Logistics Practices:

The modern day supply chains have become extremely competitive and this has led to changes in logistics practices. These changes taking place in logistics practices have also increased the pressure on the existing transportation system.

1.3.1 Shift from “Push” to “Pull” Logistics: One of the major changes that is taking place in logistics management is the shift from a “push” based logistics management system to a “pull” based logistics management system. These concepts are explained below with the help of a typical supply chain shown in Figure 1.2.

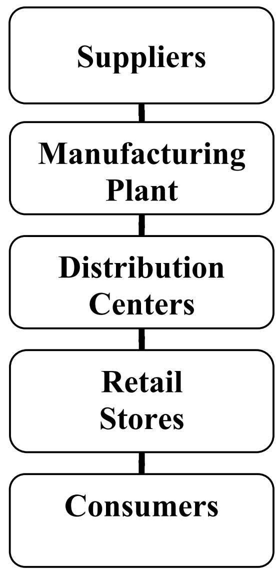


Figure 1.2: A Typical Supply Chain

1.3.2 Traditional “Push” Logistics System: “Push” logistics (or manufacture-to-supply) is an inventory based system. In this case the raw materials are pushed from the supplier (also referred to as a vendor) to a manufacturer, then finished products are pushed from a manufacturer to a distributor (also referred to as a wholesaler), who in turn pushes these products to a retailer and then the retailer fills (satisfies) the consumer’s order. Here at each level, the amount of goods to be acquired is determined based on demand forecasts. To accommodate any fluctuations in demand, an inventory is maintained at each level. The inherent disadvantage of this system is that there is a wasteful inventory stocked in the warehouses at each level and it ultimately results in an increased cost for the customer. In the above case of "push" logistics system, the carriers are only expected to deliver the goods within a reasonable amount of time since the risk of stock outs is low.

1.3.3 Modern "Pull" Logistics System: In "Pull" logistics system, the commodities are manufactured according to order, i.e., each lower level component of the supply chain is able to pull the products by placing orders depending on real time demand for the product. Unlike the "push" logistics system, this system does not depend on inventory but it relies on accurate flow of information and just-in-time delivery of goods. In this system, the logistics management is done from a holistic point of view, i.e. the overall benefit of the supply chain is considered rather than the individual components. As a result of this, inventory is reduced at all levels of the supply chain and in some cases; the need for a distributor is eliminated all together. But, the adoption of a "pull" logistics system also, comes at a cost i.e. higher risk of stock out. And as a result of this, shippers demand a highly reliable and timely delivery of shipments from carriers. On many occasions, they also want to track the exact location of their shipments in transit. Hence, this logistics system places additional demand on the transportation system.

As a result of the emergence of pull logistics, the average shipment size is getting smaller, as shippers prefer continuous replenishment using frequent shipments. Therefore, truck has become the preferred mode of transportation for many types of commodities. In particular, the demand for Less than Truck Load (LTL) carriers is increasing. The revenue and tonnage of LTL carriers are expected to grow at an annual rate of 3.0 % as compared to the 2.5 % growth of Truckload carriers, up to the year 2014 [5]. Also, the average load carried by LTL trucks is decreasing. A Bureau of

Transportation Statistics (BTS) study on carriers shows that the average LTL load has come down from 13.8 tons in 1990 to 11.9 tons in 2000 [6].

A direct offshoot of this change is the emergence of just-in-time transportation system. As, more and more industries are switching to just-in-time practices, the delivery windows are a lot tighter and the shippers expect expedited delivery of goods and in some cases they even want their goods to be delivered not earlier and not later than a certain interval of time. According to a Bureau of Labor Statistics publication, just-in-time manufacturing increased from 18 % in 1990 to 28 % in 1995 [7]. This report also states that the inventory-sales ratios are declining sharply.

1.3.4 Emergence of Electronic Commerce: E-commerce enables buying and selling goods through electronic networks (primarily through internet). With the increased usage of e-commerce, the consumers are directly interacting with suppliers, hence minimizing the need for distributors and retailers. This, in turn leads to the reduction in inventory levels and physical distribution of goods becomes a very important activity in the supply chain. Once again, because of the increased usage of e-commerce, quicker responses and faster delivery of goods are being demanded from the carriers. This is also resulting in an increased demand for LTL carriers because frequent delivery of smaller shipments is needed.

1.4 Demand Management Efforts:

In order to accommodate the growing demand for freight transportation, transportation planners are considering various innovative alternatives to accommodate the demand for freight transportation. One of the options is developing the existing infrastructure to improve intermodal transportation. Some notable developments in this regard are the Alameda Corridor in California and the proposed introduction of exclusive truck lanes linking the intermodal facilities in New York – New Jersey area [8]. Addition of new infrastructure in order to accommodate the growing demand is increasingly becoming difficult because of socio-economic and environmental constraints and sometimes even undesirable. Therefore freight planners are making efforts towards better demand management in order to make efficient use of the existing infrastructure. One of the important demand management options that is being considered is the modal diversion of freight shipments from truck to rail. Another option that is being considered by planners at various places is the introduction of differential pricing system, i.e. using different toll rates for different types of vehicles at different times of the day. This would help in mitigating the congestion during the peak periods in metropolitan areas. It is important to understand the logistics behind the freight movement to make informed public policy decisions that are aimed at effective demand management. Also, in view of the changes taking place in supply chain management practices, it is important to understand the role of transportation in the supply chains.

1.5 Problem Statement

Freight forecasting methodologies that have been used so far are typically based on the four step passenger travel demand forecasting procedure. These methodologies lack a behavioral understanding of freight movement and hence had limited applicability for freight planners. Besides, these methodologies have relied on very few aggregate data sources that lacked decision sensitive information. Hence a new methodology for forecasting regional commodity flows that captures both the spatial and behavioral elements of freight movement is required.

1.6 Purpose and Scope

The purpose of this study is to demonstrate the applicability of a supply chain based modeling methodology for regional freight forecasting. The methodology consists of the following two steps: 1) Obtaining O-D Flows by tracing the supply chains 2) Modeling the mode choice decision process of shippers. The rationale behind such a methodology is to capture both the spatial and behavioral elements of the supply chains. A supply chain based methodology is used because freight movements are a result of supply chain practices of individual firms. The possibility of using additional data sources that might be useful in providing a better understanding of the underlying behavior behind freight movement is also explored in this study.

The methodology is not applied to all the commodities; it is demonstrated using only a few commodities. However, this methodology is transferable across all commodities. Individual shipment level information is not collected due to confidentiality concerns and information is collected only at a firm level. Though firm level information is collected for the study; the responses of individual firms are not published to protect confidentiality.

Chapter 2

Literature Review

2.1 Introduction

The efficiency of a national economy is inter-dependent on the efficiency of the logistics system in a country. In efficient economies, the total logistics costs are about 9% of the cost of the product. On the other hand the total logistics cost can be as high as 30% in some of the developing countries [9]. Some of the important barriers to efficient supply chains include poor transportation infrastructure, non competitive markets, lack of market information and improper transportation regulation. A literature review was conducted to understand the trends in logistics practices, their effect on transportation, the different sources of freight data available and the models that are being used in freight transportation planning.

2.2 Problems with Existing Freight Planning Methodologies

An important drawback of the existing literature in freight planning is the missing link between the freight planning practices in the public sector and the supply chain management practices in the private sector. Though these two processes are highly interrelated, the literature that links these two processes is scant.

Individual firms take transportation decisions as a part of the larger process of optimizing the total supply chain performance. In other words, the firms make their transportation decisions with the objective of minimizing the supply chain costs rather than minimizing the transportation costs. Hence the freight demand models used for transportation planning should focus on capturing the interactions between the transportation variables and other supply chain variables that affect the transportation decisions of these firms. The total supply chain costs can be viewed as two components: 1) The tangible logistics costs 2) The intangible service related costs. The existing freight demand models have failed to take into account all the important variables of these two components; and the following sections describe the costs associated with these.

2.3 Logistics Costs:

The important components of total logistics cost are transportation costs, warehousing costs, order entry/customer service costs, administrative costs and inventory carrying costs. An annual study of logistics costs by Herbert Davis Company provides the following breakdown of the total logistics costs [10].

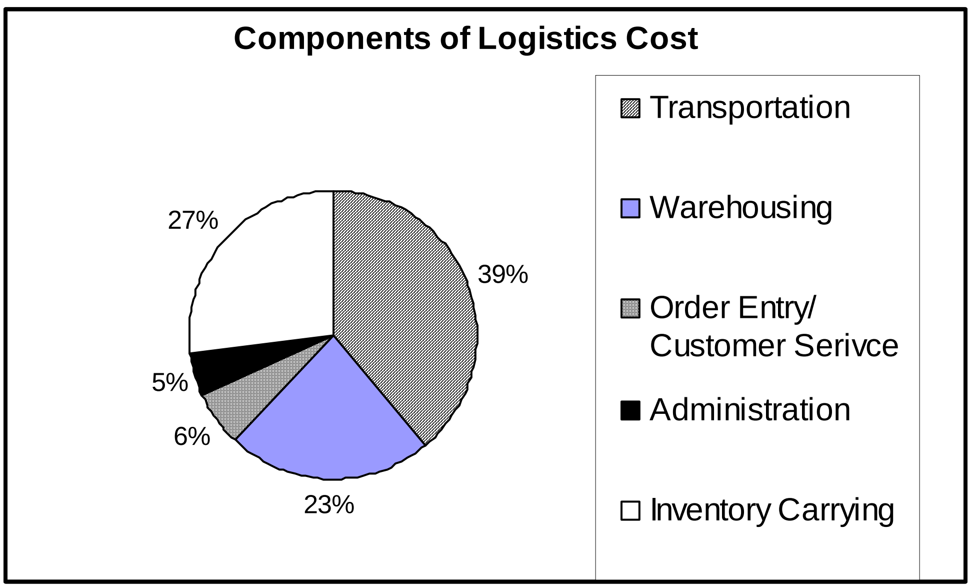


Figure 2.1: Components of Logistics Cost

Source: Reference [10]

About 61 percent of the total logistics costs comprise of non-transportation related logistics costs. However, the inventory carrying costs and warehousing costs are highly interdependent on the efficiency of the transportation system.

2.4 Trends in Logistics Costs

The logistics costs have ranged between 5-10 percent of the sales revenue (or product cost) since the 1960s. The logistics costs were at 10 percent of the sales revenue in early 1960s before firms realized physical distribution was an important operation that required special attention. After the firms have realized the importance of physical distribution and started focusing on eliminating inefficiencies in their distribution costs the logistics costs came down to about 5 percent of the sales revenue. Fuel crisis and high inflation resulted in an increase in logistics costs during the 1970s. In the early 1980s the logistics costs were close to 10 percent of the sales revenue. In 1980s deregulation of railways coupled with improvements in logistics practices resulted in a decline in logistics costs. By early 1990s logistics costs were at 7 percent of the sales revenue and they remained steady through out the 1990s. In 2001 a sharp recession in the economy resulted in an increase in logistics costs to 8.5 percent. Since then the logistics costs have remained steady between 7 to 8 percent. There have been some major changes in supply chain practices since 2001. During this period many firms have started sourcing from overseas locations. This has made their supply chains global and this had an increasing effect on the logistics costs. On the other hand, the increased use of information technology in logistics management had improved the efficiency of supply chains and this had a balancing effect on the total logistics cost [10].

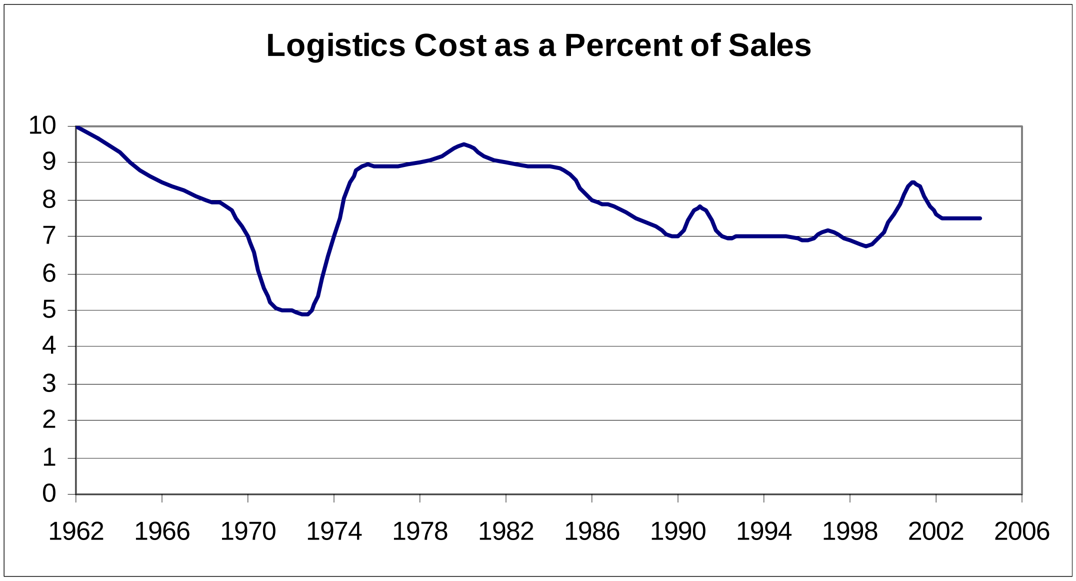


Figure 2.2: Annual Trend in Logistics Cost to Sales Ratio

Source: Reference [10]

2.5 Factors Affecting Total Logistics Cost to Sales Ratio

The study [10] conducted by Herbert Davis Company also shows that smaller companies incur higher logistics costs as compared to larger companies. The averages logistics costs as a percentage of sales is 11 percent for companies with an annual sale of less that 200 Million dollars as compared to the average 5.4 percent for companies with annual sales greater that 1.25 Billion dollars.

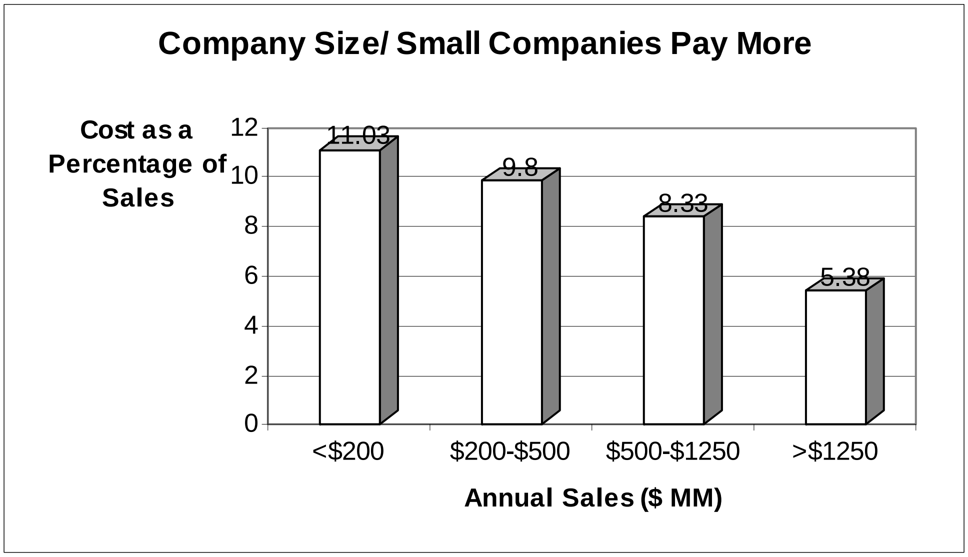


Figure 2.3: Variation of Logistics Cost to Sales Ratio with Company Size

Source: Reference [10]

Another factor affecting the logistics cost to sales revenue ratio is the value of the product. The manufacturers of food products whose product value is about 1.5 dollars per pound spend about 10 percent of their revenue on logistics costs. On the other hand manufacturers of high valued products like electronic equipment with a value greater than 15 dollars per pound spend about 3 to 4 percent of their revenue on logistics costs. However, the actual logistics costs involved with high valued products are higher as compared to the actual logistics costs involved with low valued products [10].

The product cycle time affects the inventory carrying costs as there is a capital cost associated with commodities that are held up in the inventory. An average cycle time of 8.4 working days was reported in 2004 for in-stock items. However, a disadvantage with the findings of this study is that all the findings were reported as an average of all the commodities. The commodity category wise findings were made available only to the participants of the study [10].

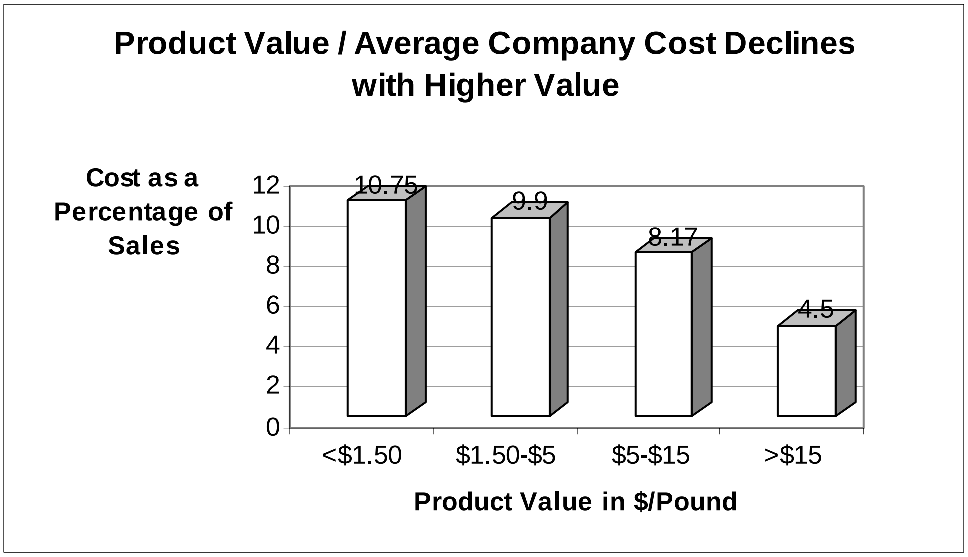


Figure 2.4: Variation of Logistics Cost to Sales Ratio with Product Value

Source: Reference [10]

2.6 Service Related Costs

These are the costs that are incurred because of the lost sales opportunities or because of the lack of availability of the right product at the right time at the right place. General Motors estimates that about 10 percent of sales are lost because the car is not available [11]. Some of the factors that affect these costs are: reliability of transportation, product characteristics like perishability, lead time etc. It is difficult to calculate the cost associated with these factors directly.

2.7 Comparison of Truck and Rail

Researchers from Cap Gemini, Ernst & Young, Georgia Southern University, Logistics Management and the University of Tennessee conducted a study in the year 2003 to evaluate the performance of different trucking modes and rail [12]. This study obtained responses from one hundred and eighty eight shippers representing all major industrial sectors on five service dimensions namely: On-time delivery ratio, Equipment availability, Billing error rate, Freight loss and damage and Turndown ratio. The graphs showing the performance of different modes on service dimensions relevant to the current research project are shown below. Shipments by truck have shown an on-time delivery of about 95 percent as compared to the on-time delivery of 84 percent for rail. The average freight loss and damage rates were comparable at about 1.2 percent for both truck and rail. The average equipment availability for the trucking modes was about 95 percent as compared to 90 percent for rail.

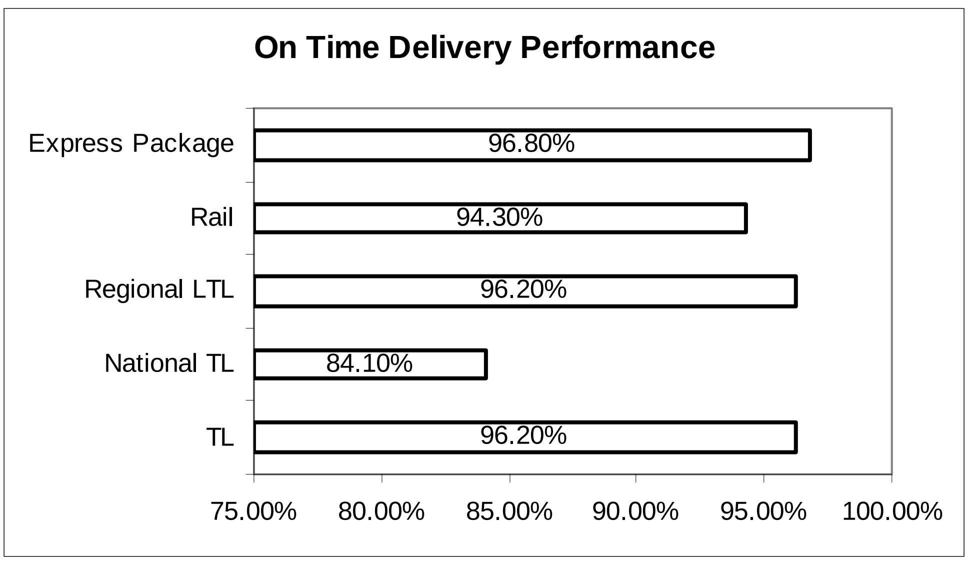


Figure 2.5: Modal Comparison of On-Time Delivery Performance

Source: Reference [12]

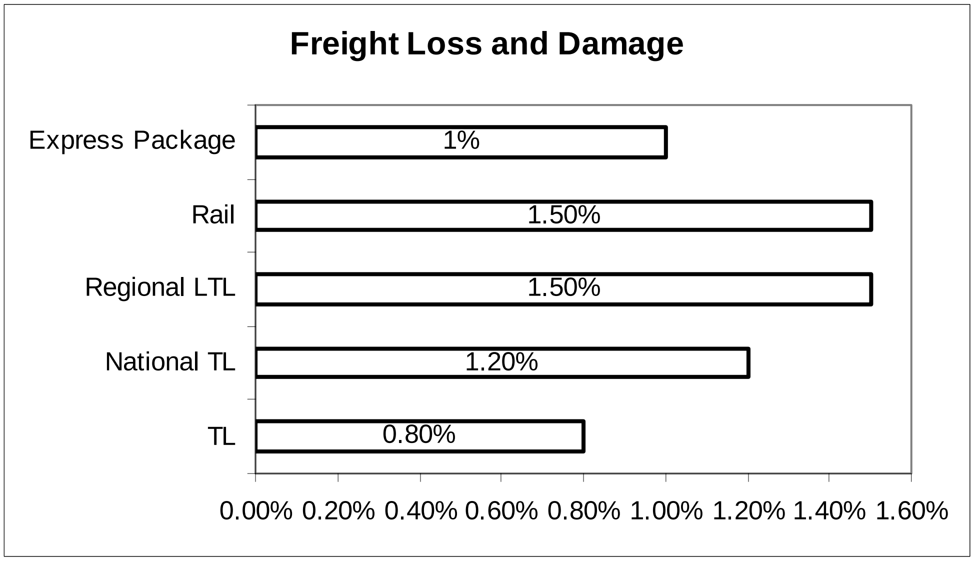


Figure 2.6: Modal Comparison of Freight Loss and Damage

Source: Reference [12]

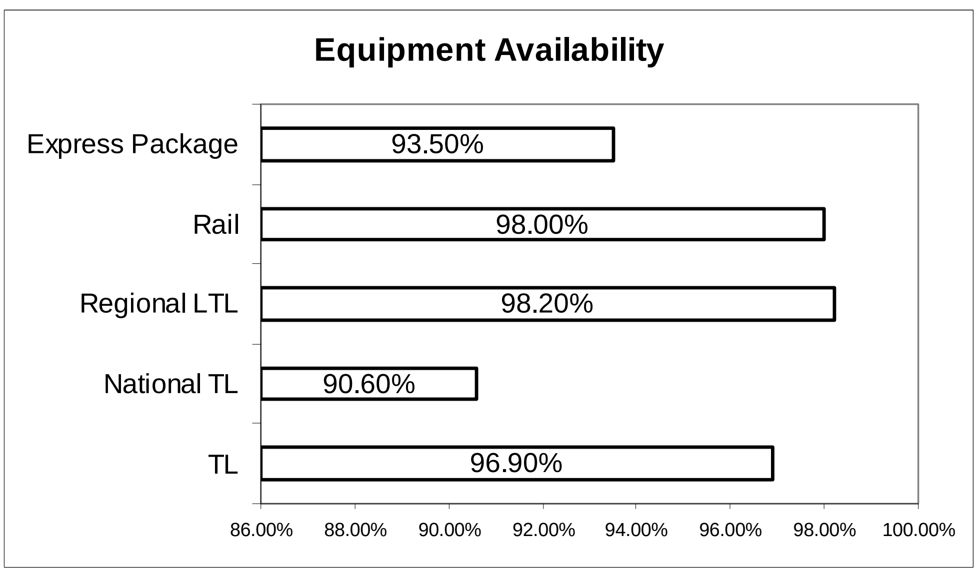


Figure 2.7 Modal Comparison of Equipment Availability

Source: Reference [12]

2.8 Freight Transportation Planning Models

In spite of the growing importance of integrating freight movement into the transportation planning process, the research in freight modeling is lagging behind the research in passenger modeling. One of the major reasons cited for this has been the lack of availability of publicly available freight data [13]. Even the few sources of freight data that are publicly available are published aggregately to protect the identity of individual shippers. Another reason for the lack of sufficient advances in freight modeling is due to the fact that freight modeling is inherently more complicated than passenger modeling. This is because of the following reasons:

- There is a large variation in freight shipment characteristics due to differences in shipment size, value, perishability etc.
- Freight decision making involves a complex interaction between the shipper, receiver and carrier and none of them has complete information or decision making power.
- Freight transport prices are usually negotiated as a long term contract and they are not uniform for all the shippers.

2.8.1 Trip Generation and Distribution Modeling

Most of the freight demand models that have been developed have closely parallel the four step passenger planning process which involves modeling of trip generation, trip distribution, mode choice and traffic assignment. In trip generation modeling, the trip productions and attractions are usually modeled by regression on socio-economic factors like population, employment, per capita income and area [3, 14]. This approach is justifiable in case of passenger trip generation modeling because the above socio-economic factors are explanatory variables for passenger trips. But in case of freight trip production modeling, these socio economic factors are not explanatory as freight trip productions only depend on the presence of the particular industries and their output. Even in case of freight trip attraction these socio-economic variables are not explanatory for Business to Business freight attractions and they are only explanatory for Business to Retailer/ Consumer trips.

Gravity models have been commonly used for trip distribution modeling. However, the problem with the application of gravity models for freight trip distribution modeling is that the friction factors are different for different modes and modal split of the trip generation needs to be known beforehand. This leaves the modeler in a paradoxical situation as modal split is usually the step following trip distribution.

2.8.2 Mode Choice Modeling

The use of trip generation and distribution models is reasonably well developed in freight forecasting [14, 15]. However, the modeling of mode choice has been the most difficult step for most practitioners and the research into this step is still too elementary to be included in the freight forecasting models [16]. Several studies have reported a failure or difficulties in developing mode choice models [14, 17]. Some of the difficulties in developing mode choice models are obscurity in the identification of mode choice decision maker(s), lack of proper understanding of the mode choice decision process and lack of availability of reliable disaggregate data. The potential for the development of discrete choice models using aggregate data has not been explored.

Disaggregate demand models have been generally used for mode choice modeling. These models have been classified as behavioral models and inventory models [18]. Behavioral models like logit and probit use the theory of utility maximization in which the mode with the maximum utility is chosen by the shipper. Inventory based models take the perspective of a firm's inventory manager and attempt to link the mode choice and production decisions of a firm. However, the need for detailed firm level data makes the implementation of inventory models impractical for planning purposes. Abdelwaheb and Sargious have developed a switching simultaneous equation model that estimates the mode choice and shipment size simultaneously [19]. They argue that using a single equation model for estimating the mode choice introduces a potential bias. However, in reality most firms do not simultaneously determine the mode choice and shipment size. The mode choice decision is usually a long term decision as the contracts between shippers and carriers last between three to five years [20]. The shipment size is a short term decision process which can be a daily decision for some of the firms.

Since the availability of reliable freight data at a disaggregate level is difficult, the use of some unconventional methods for mode choice modeling has been explored in the recent past. Sen, Pozzi and Bhat have used the Delphi Technique for mode choice analysis [21]. The expert panel that participated in this study consisted of Metropolitan Planning Organization (MPO) planners, state planners and port, truck and rail representatives. However my opinion is that an expert panel that consists of logistics managers from shipping firms who are the actual decision makers would have been more representative for this kind of study.

Another innovative approach being used in freight mode choice analysis is the use of stated preference data. Daniels, Marcucci and Rotaris have used the stated preference data collected from logistics managers to model the choice of mode [22]. Adaptive Conjoint Analysis (ACA) software was used in this study to collect the preferences among freight service attributes from the logistics managers. Some of the advantages in using stated preference data are: 1) It is relatively easier to obtain stated preference data as it need not be confidential 2) It allows the modeler to control the variability in attributes 3) It provides the ability to model future scenarios. The disadvantage with the use of stated preference data is that the choices are hypothetical [23].

2.9 Summary

The existing literature relevant to freight transportation planning lacks an understanding of the logistics behind the movement of freight. This is because the current literature exists as two distinct entities: one part of the literature, present in business literature, deals with the logistics and supply chain management practices of individual firms and the other part, present in transportation literature, deals with the freight models used by Transportation planners. Only a few studies [24, 25] have attempted to include the logistics processes in freight transportation models. The methodology presented in the next chapter integrates private sector supply chain practices into public sector transportation planning.

Chapter 3

Methodology

3.1 Need for a Supply Chain Based Modeling Methodology

As the demand for freight transportation is growing at a rate greater than what the present transportation infrastructure can handle, new measures of effective demand management are required. The most significant means of demand management that are being considered are modal diversions from truck to rail and the introduction of differential pricing systems on highways. In order to decide upon the appropriate measures of demand management and to estimate the effectiveness of these measures, understanding the logistical characteristics of the freight shipments is necessary.

For example; in case of planning modal diversion measures, it is important to understand all the important links of a supply chain as the logistics that govern the movement of goods in each link of the supply chain are different. Supply chain links can be classified as Business-to-Business links and Business to Customer (or Retailer) links. The later category of links is more time sensitive and it requires frequent delivery of smaller shipments. Also in Business to Customer links, the final customer doesn’t act under a contract with the retailer. Whereas the former category of links would be less time sensitive and the size of the shipments in this case will be larger. Hence, if one were considering the potential for highway freight traffic diversion, it would be helpful to consider only the shipments belonging to Business-to-Business links of a supply chain. Also with in the Business-to-Business shipments, the types of commodities that have the potential to be diverted from truck to rail have to be identified first. Because of the logistical characteristics like time sensitivity, risk of damage, perishability etc., some of the commodities do not provide for a choice of mode.

Similarly while considering differential road pricing, a thorough understanding of the logistical characteristics of the commodity movements is required. The impact of differential pricing on different industries needs to be considered, because some industries must ship commodities during peak periods due to constraints on customer service, production schedules etc. Also, the value of transportation time for various commodities and the amount of tolls the firms involved in the supply chains are willing to pay on tolled facilities needs to be understood.

As logistics is the driving force behind the transportation decisions of any shipper, it would be appropriate for a modeling methodology to be based on the logistical characteristics of the commodities.

3.2 Problems with the Conventional Data Sources

The Commodity Flow Survey (CFS) and the TRANSEARCH database are two of the most popular sources of data that are used by freight planners. However, these two data sources suffer from several limitations. Though the CFS is collected at individual shipment level, the final results of the CFS are only at the state level in order to avoid disclosure of the operations of any individual firm or establishment. Moreover the flows are provided at the two-digit Standard Classification of Transported Good (SCTG) level. The TRANSEARCH database attempts to address some of the deficiencies of the CFS by providing freight flows at a County level and at a 4-digit level of Standard Transportation Commodity Codes (STCC) Classification. The major concern with the TRANSEARCH database is that since this database is proprietary, very little information is available about the construction of this database and the accuracy of the data. However, the TRANSEARCH database is widely used by various organizations for freight planning.

3.3 Proposed Modeling Methodology

A two step methodology which makes use of additional data sources other than the conventional data sources like the Commodity Flow Survey and TRANSEARCH database is suggested for freight modeling. Apart from linking the private sector supply chain practices to public sector transportation planning, this methodology attempts to overcome the limitations of current freight trip generation, trip distribution and mode choice models discussed in the previous chapter. In particular, the issue of developing a mode choice model in the absence of disaggregate data is addressed in this methodology. The two steps in the methodology are described below:

- Obtaining O-D Flows by Tracing the Supply Chains: On tracing the supply chains of major business units in a region, the origins and destinations of the flows can be located. This could be used in combination with market share analysis and the sales volume from an individual firm's annual report to obtain the O-D flows. This is equivalent to the trip generation and distribution steps of the 4-step planning process. This step is also useful in understanding the accuracy of TRANSEARCH database.
- Mode Choice Analysis: The logistical needs and constraints of a shipper determine the choice of mode. Therefore, the mode choice analysis that accounts for the logistical variables would be appropriate. The important supply chain

variables that affect the choice of mode need to be identified by reviewing the supply chain literature and/or by consulting the actual supply chain decision makers of individual firms. After identifying the important supply chain variables mode choice analysis can be performed using an analytical method or by developing more rigorous empirical models based on observations of choice that have already been made. An analytical method can be used after collecting the relative importance of the supply chain variables by surveying a sample of shippers. The empirical model can be developed using disaggregate shipment level data or using aggregate data at a county level.

3.4 Illustration of the Methodology

A brief description of the methodology is provided below. A detailed description of its application is provided in the subsequent chapters. This two step modeling methodology combines the information available from multiple data sources like TRANSEARCH, InfoUSA, case studies available from supply chain literature and publicly available data from individual firms. This data has been augmented with data obtained from a confidential survey of shippers.

"Motor Vehicles" classified as STCC 3711 was used to demonstrate the Step 1 of the above methodology. InfoUSA[1] database was used to locate the Motor Vehicle manufacturers in Virginia. This database has shown that Volvo's manufacturing plant in

[1] InfoUSA database provides commodity wise listing of all Businesses in any geographic region within the United States. It also provides information like number of employees, annual sales volume etc. for each firm.

Pulaski County is the only Motor Vehicle manufacturing plant in Virginia. Hence, a Case Study on Volvo Trucks was used to obtain the O-D flows from Pulaski County.

This case study, which is described in Chapter 4, has been prepared by conducting an exhaustive search of the information available regarding Volvo like the locations of their suppliers, dealers, number of units manufactured per year and the logistics management for Volvo. Volvo has been selected for a case study because of the location of the manufacturing plant in Virginia. As the TRANSEARCH database provides county level commodity flows at a four digit STCC level for the state of Virginia, this case facilitated a direct comparison between the commodity flows from Pulaski County as provided by TRANSEARCH database and the expected commodity flows based on the Volvo’s annual truck sales and dealership locations. This case study has been useful in the verification of TRANSEARCH database.

For mode choice analysis, the commodities Motor Vehicles (STCC 3711), Fiber, Paper or Pulp Board (STCC 2631) and Meat Products (STCC 2013) were considered. The commodities Motor Vehicles (STCC 3711) and Fiber, Paper or Pulp Board (STCC 2631) were used because for these commodities the truck and rail were both viable alternatives. The use of commodities STCC 3711 (a relatively high valued commodity) and STCC 2631 (a relatively low valued commodity) has ensured that the mode choice analysis was done on two commodities of contrasting commodity values. The commodity Meat Products (STCC 2013) was also included because it is a perishable commodity and it is expected to have very different logistical characteristics as compared to the non-perishable commodities.

The database InfoUSA is used to identify the shippers of Motor Vehicles (STCC 3711), Fiber, Paper or Pulp Board (STCC 2631) and Meat Products (STCC 2013) manufacturers. A confidential survey was sent out to the shippers identified above. This survey obtained information about the relative of importance of attributes like transportation costs, logistics costs, travel time, travel time reliability, risk of loss or damage etc. It also obtains the perceived values of these attributes over distances of 200, 500 and 1000 miles for truck and rail modes. The results of this survey are summarized and an analysis of the important factors affecting the choice of mode for each of the above three types of commodities is provided in chapter five.

Rigorous empirical modeling techniques that can be used for modeling the choice of mode are presented in chapter six. Two data sets pertaining to outbound shipments from Arlington and King William counties are used for model calibration and testing respectively. These data sets have been extracted from the TRANSEARCH database. This data has been supplemented with data obtained from the survey of shippers. Empirical modeling techniques like Logit models, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Classification Trees have been used and their performances compared on the above data sets.

Chapter 4

Volvo Trucks Case Study

4.1 Corporate History

Volvo group had started its operation in Sweden in the year 1927. It started manufacturing trucks in the year 1928 [26]. Volvo currently manufactures Trucks, Cars, Buses, Construction Equipment, Industrial Engines and Aircraft Engines [27]. Volvo entered the U.S. truck market in 1959. In the year 2001 Volvo acquired Mack Trucks, which is one of the largest manufacturers of heavy-duty trucks in the U.S.

Volvo manufactures Class 8 trucks in the U.S. The Volvo’s New River Valley plant in Virginia’s Pulaski County is the only manufacturing plant for Volvo trucks in North America. This plant has been in existence since 1984. Prior to Volvo’s acquisition of Mack trucks, the production facility of Mack trucks was located in Winnsboro, South Carolina. This production facility was closed in November 2002 and the entire production was shifted to New River Valley plant by May 2003 [28].

4.2 The Volvo Supply Chain

4.2.1 Suppliers

The Volvo truck manufacturing plant in Pulaski County, Virginia obtains various parts from suppliers all over the world. Two of the important suppliers have been identified as Volvo Powertrain and ArvinMeritor.

4.2.1.1 Volvo Powertrain: The Volvo Powertrain, located in Hagerstown, Maryland, provides the entire Volvo group of trucks with diesel engines, transmissions and axles. It either manufactures or purchases these components. The Hagerstown plant has started supplying diesel engines for the entire Volvo group of trucks manufactured at the New River Valley plant in 2003. Prior to the year 2003, the Hagerstown plant used to supply engines for Mack trucks and the powertrain plant in Skovde, Sweden used to supply the engines for Volvo trucks.

4.2.1.2 ArvinMeritor: ArvinMeritor Inc., a global supplier of a broad range of components to the motor vehicle industry, was formed in the year 2000 by the merger of Meritor Automotive Inc. and Arvin Industries Inc. ArvinMeritor supplies braking systems to the Volvo trucks manufacturing plant. ArvinMeritor's manufacturing facilities in Manning, South Carolina and Tilbury, Ontario provide the braking systems for the New River Valley plant [29].

4.2.2 Dealer Network

Volvo trucks has a dealer network across all the 50 states and Washington D.C. in the U.S. The number of dealers in each state is shown in Table 4.1 [30].

4.2.3 Supply Chain Management

Logistics: Volvo logistics provides the logistics capability for the truck manufacturing plant. It takes care of the entire inbound, outbound and in-house logistics requirement for the New River Valley Manufacturing plant [31].

Purchasing: Volvo 3P provides product planning, product development, purchasing and product range management for Volvo trucks.

Table 4.1: Number of Tons of STCC 3711 Shipped from Virginia's Pulaski County to Each State in the U.S. for the year 2003 [30, 32].

State	Dealers	No. of Trucks	Tonnage	State	Dealers	No. of Trucks	Tonnage
Alabama	6	584	3796	**Montana**	3	292	1898
Alaska	0	0	0	**Nebraska**	1	97	631
Arizona	3	292	1898	**Nevada**	1	97	631
Arkansas	5	487	3166	**New Hampshire**	1	97	631
California	9	877	5701	**New Jersey**	8	779	5064
Colorado	2	195	1268	**New Mexico**	2	195	1268
Connecticut	2	195	1268	**New York**	12	1169	7599
Delaware	2	195	1268	**North Carolina**	8	779	5064
D.C.	1	97	631	**North Dakota**	3	292	1898
Florida	7	682	4433	**Ohio**	12	1169	7599
Georgia	6	584	3796	**Oklahoma**	2	195	1268
Hawaii	1	97	631	**Oregon**	5	487	3166
Idaho	1	97	631	**Pennsylvania**	18	1753	11395
Illinois	10	974	6331	**Rhode island**	1	97	631
Indiana	9	877	5701	**South Carolina**	4	390	2535
Iowa	5	487	3166	**South Dakota**	2	195	1268
Kansas	2	195	1268	**Tennessee**	5	487	3166
Kentucky	4	390	2535	**Texas**	18	1753	11395
Louisiana	4	390	2535	**Utah**	2	195	1268
Maine	3	292	1898	**Vermont**	2	195	1268
Maryland	4	390	2535	**Virginia**	8	779	5064
Massachusetts	3	292	1898	**Washington**	4	390	2535
Michigan	5	487	3166	**West Virginia**	6	584	3796
Minnesota	5	487	3166	**Wisconsin**	6	584	3796
Mississippi	6	584	3796	**Wyoming**	1	97	631
Missouri	6	584	3796	**Puerto Rico**	1	97	631
				Total	247	24055	156358

4.3 Volvo Sales and Market Share

Table 4.2 provides the total annual sales and market share for Volvo trucks in between the years 1998-2003 along with the individual brands.

Table 4.2: Annual Market Share and Sales Information for Volvo Trucks [32]

Year	**Volvo Sales**	**Market Share**	**Mack Sales**	**Market Share**	**Total Sales**	**Total Market Share**
1998	24060	9.7	N.A.	N.A.	24060	9.7
1999	28177	10.7	N.A.	N.A.	28177	10.7
2000	22565	10.7	N.A.	N.A.	22565	10.7
2001	13964	10	20351	14.6	34315	24.6
2002	11025	7.5	20482	13.6	31507	21.1
2003	13711	9.7	15146	10.7	28857	20.4

4.4 Comparison of Flows with TRANSEARCH Data

Volvo's only truck manufacturing plant in North America is located in Virginia's Pulaski County. It supplies Trucks to all its dealers in the 50 states of U.S. However, the TRANSEARCH database shows flows corresponding to Commodity STCC 3711 (Motor Vehicles) only into Lexington (KY), Chicago (IL), Tennessee and the East South Central Census Division consisting of the states of Kentucky, Mississippi, Alabama and Tennessee. There are no flows to any other geographic region. This shows that the information about these flows from Pulaski County is not accurate. Now using the above sales information, a more accurate estimate of the outbound flows from Pulaski County for the commodity STCC 3711 for the year 2003 was made. These flows are shown in Table 4.1. A comparison of the estimated commodity flows originating from the Pulaski County for the year 1998 and the flows shown in the TRANSEARCH database for the year 1998 is shown in Table 4.3. It was assumed that each empty truck weighs 6.5 tons.

As the information about the number of trucks sold by each dealer is not publicly available, it was assumed that all the dealers would be selling an equal number of trucks.

Table 4.3: Comparison of Estimated Flows with TRANSEARCH Flows for 1998

Origin	Destination	STCC	TRANSEARCH	Dealers	Estimated
Pulaski County, VA	Lexington, KY	3711	1937	1	631
Pulaski County, VA	EAST SOUTH CENTRAL	3711	52608	15	9465
Pulaski County, VA	Chicago, IL	3711	25210	3	1893
Pulaski County, VA	Tennessee (rest of), TN	3711	28684	5	3155
			108439		15144

4.5 Summary

This case study is an illustration of the changes taking place in supply chain practices. There is an increasing trend towards mergers and acquisitions and the supply chains are becoming global. In auto industry some of the examples of mergers and acquisitions other than Volvo and Mack are Daimler and Chrysler, Mitsubishi and Fuso. These mergers and acquisitions are helping these firms to maintain localized sources of supply. The powertrain facility in Hagerstown, Maryland is an example of localized supply. This helps Volvo in having more reliable lead times. Another example of a trend towards more localized sources of supply is Wal-Mart, which is adding about 40 distribution centers every year.

This case study demonstrates how commodity wise O-D Flows can be obtained at a county level. The commodity flows obtained in this case study have accurate Origins and Destinations as compared to TRANSEARCH database. The commodity flows are also more accurate in terms of magnitude because they are based on actual sales volume data. As there is very little information available about the accuracy of commodity flows, this method can be used to supplement the TRANSEARCH database. This case study shows that it is not possible in all cases to maintain confidentiality of the commodity flows when the O-D flows are published at a county level for four digit STCC codes.

The demonstrated method of obtaining the O-D Flows is data intensive and it can be tedious to obtain the required data. It might not be possible to completely trace the supply chain for any company without their involvement in the study. If the purpose of the study is public sector planning, it may be difficult to obtain such collaboration. Even if it is not possible to trace the entire supply chains of all the companies, this method can be used to estimate the amount of each commodity produced (supply) and needed (demand) by each county. Then the Supply and Demand for each commodity can be balanced by a method similar to the gravity model to obtain the O-D flows.

CHAPTER 5

Study of Factors Affecting the Choice of Mode

The research in the freight mode choice modeling is still elementary to be included in the freight forecasting techniques [16]. The models that have been developed so far do not adequately capture the logistics behind the movement of freight. Different commodity groups have different logistical characteristics and consequently the factors that influence the choice of mode are different. This chapter identifies the important supply chain variables that affect the choice of mode and attempts to find the relative importance of these variables for three different commodity groups in determining the choice of mode.

5.1 Identification of Supply Chain Variables to be Studied

Based on literature review regarding the logistics processes in different firms and interviews with logistics managers of three major retail firms, the following set of decision variables have been identified as those that could potentially influence the transportation decision process of a firm [20, 33].

5.1.1 Shipper characteristics:

a) Annual volume of shipments (in weight)

b) Indicator of the size of the firm (e.g.: annual sales/ number of employees)

c) Average shipment distance

d) Number of O-D points served

5.1.2 Commodity characteristics:

a) Value of the commodity (in dollars per ton)

b) Density of the commodity

c) Shelf life (if the product is a perishable)

5.1.3 Logistic characteristics:

a) Total logistics cost

Total logistics cost includes the following costs:

i) Order processing costs

ii) Product handling and storage costs

iii) Transportation costs

iv) Capital costs of goods in inventory and transit

v) Stock out costs in case of late shipments

b) Total cycle time (storage time+ transportation time)

c) Shipment size (Weight/Volume)

d) Shipment frequency

e) The position of the firm in the supply chain (i.e. Supplier/Manufacturer/ Distributor/Retailer or if it is a combination of these functions)

f) Maximum acceptable delay

5.1.4 Modal characteristics:

a) Rate per mile

b) Trip time

c) Percentage of loss and damage

d) Percentage of on-time delivery

Though the above variables are important in determining the choice of mode, many of them are likely to show strong correlations to each other and they may not enter the final mode choice model. The data regarding the above variables was obtained with the help of a questionnaire to be completed by the shippers. The collected data was limited by the availability of time and resources.

5.2 Design of the Questionnaire

A comprehensive questionnaire intended for the logistics managers of shipping firms was designed for this study. This questionnaire was intended to collect disaggregate shipment level data from individual shippers. The availability of disaggregate shipment level data would be very helpful in developing a versatile mode choice model. A copy of this comprehensive questionnaire is provided in Appendix A. However, the comprehensive questionnaire was not used because of the following reasons: 1) The time required to complete the survey was expected to be more than 30 minutes. 2) The shippers would be reluctant to provide individual shipment level information.

In order to reduce the survey burden and to improve the response rate from the shippers, a more concise questionnaire intended to collect the stated relative preferences among a selective set of attributes was designed. The concise questionnaire was actually used to collect the data. This questionnaire also collected the values of travel time, on-time performance, transportation cost as a percentage of shipment value and other logistics cost as percentage of shipment value for truck and rail over distances of 200 miles, 500 miles and 1000 miles from the shippers. A copy of this concise questionnaire that was used in collecting the data from the shippers is provided in Appendix B.

5.3 Recipients of the Survey

This survey was sent out to manufacturers of commodities Motor Vehicle (STCC 3711), Fiber, Paper or Pulp Board (STCC 2631) and Meat Products (STCC 2013). The first two commodities were selected because they have a significant share for both truck and rail modes and their commodity values per ton differ significantly and are expected to show different time sensitivity. The later product was included to obtain responses from perishable product manufacturers as perishable product are expected to show different logistical properties as compared to non-perishables like motor vehicles and fiber, paper and pulp boards. The manufacturers of the commodities STCC 3711, STCC 2631 and STCC 2013 were identified using the InfoUSA database.

Senior logistics executives of these manufacturing firms usually holding the designations “Vice President of Logistics” or “Director of Logistics” were identified as the potential respondents to the survey. Senior logistics executives were contacted because they are the persons involved in major transportation related decisions like mode choice and they usually have the authority to respond to the survey unless prohibited by a firm wide policy. The names and the contact information of these executives were obtained from internet searches and by making phone calls to these manufacturers. These executives were contacted and their preference to receive the survey electronically or via fax was collected. After this the survey was sent out to these executives and their responses collected. This survey was sent out to 40 logistics executives and 14 responses were obtained.

5.4 Summary of Survey Responses

The responses obtained from the shippers are summarized in Tables 5.1 to 5.3 and these responses are graphically representations using a series of bar charts that display the relative importance of the attributes that affect the choice of mode.

Table 5.1: Relative Weights of Attributes for All Shippers, Shippers Using Only Truck and Shippers Using Both Truck and Rail

Factor	Relative Weights		
	All	**Truck**	**Both**
Travel Time	23.0	28.6	17.4
On-time Performance	28.6	33.6	23.7
Transportation Costs	19.7	10.7	28.7
Other Logistics Costs	4.7	2.1	7.2
Ability to Track	6.0	8.6	3.5
Special Handling Equipment	4.7	5.0	4.3
Risk of Loss or Damage	5.1	3.6	6.6
Geographic Coverage	7.4	7.9	7.0
Others	0.7	0.0	1.4
Total	100	100	100

Table 5.2: Relative Weights of Attributes for All Shippers by Commodity Type

Factor	Relative Weights		
	STCC 2631	**STCC 3711**	**STCC 2013**
Travel Time	31.3	15.9	23.3
On-time Performance	19.2	25.9	50.0
Transportation Costs	21.2	24.2	8.3
Other Logistics Costs	2.5	8.9	0.0
Ability to Track	5.5	6.2	6.7
special Handling Equipment	2.7	4.5	8.3
Risk of Loss or Damage	4.8	7.0	1.7
Geographic Coverage	12.8	5.8	1.7
Others	0.0	1.7	0.0
Total	100.0	100.0	100.0

Table 5.3: Relative Weights of Attributes by Commodity Type for Shippers Using Only Truck and Shippers Using Both Truck and Rail

Factor	Relative Weights					
	2631		3711		2013	
	Truck	Both	Truck	Both	Truck	Both
Travel Time	47.5	20.6	17.5	15.0	23.3	N.A.
On-time Performance	12.5	23.6	30.0	23.8	50.0	N.A.
Transportation Costs	12.5	26.9	12.5	30.0	8.3	N.A.
Other Logistics Costs	0.0	4.2	7.5	9.5	0.0	N.A.
Ability to Track	7.5	4.2	12.5	4.0	6.7	N.A.
Special Handling Equipment	0.0	4.4	5.0	4.3	8.3	N.A.
Risk of Loss or Damage	2.5	6.4	7.5	6.8	1.7	N.A.
Geographic Coverage	17.5	9.7	7.5	6.7	1.7	N.A.
Others	0.0	0.0	0.0	2.5	0.0	N.A.
Total	100.0	100.0	100.0	99.8	100.0	N.A.

Figure 5.1 shows the average weights (on a scale of 100) assigned to each of the attributes among all the shippers. Travel time, on-time performance and transportation costs are the major factors influencing the choice of mode accounting for about 70 percent of the total weight. Figures 5.2 and 5.3 show the relative preferences among attributes for shippers that use truck only and for shippers that use both truck and rail respectively. For shippers that use both truck and rail; total logistics cost is the most important factor along with travel time and on-time performance. For shippers that use only truck; travel time and on-time performance are the only important factors.

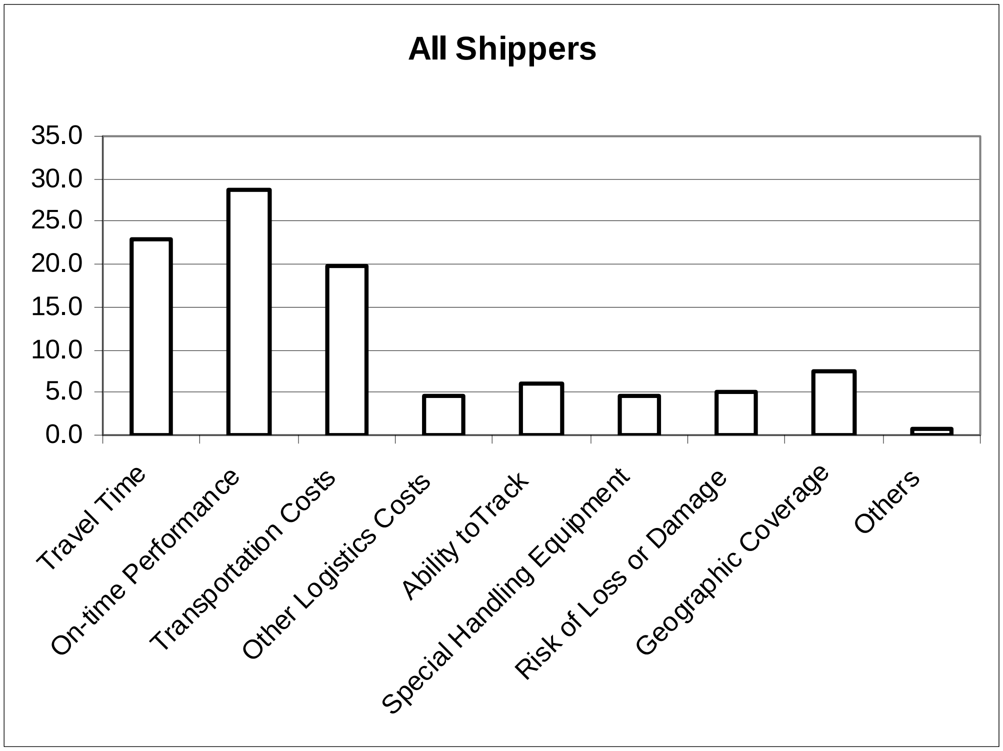


Figure 5.1: Average Relative Weights of Attributes for All the Shippers

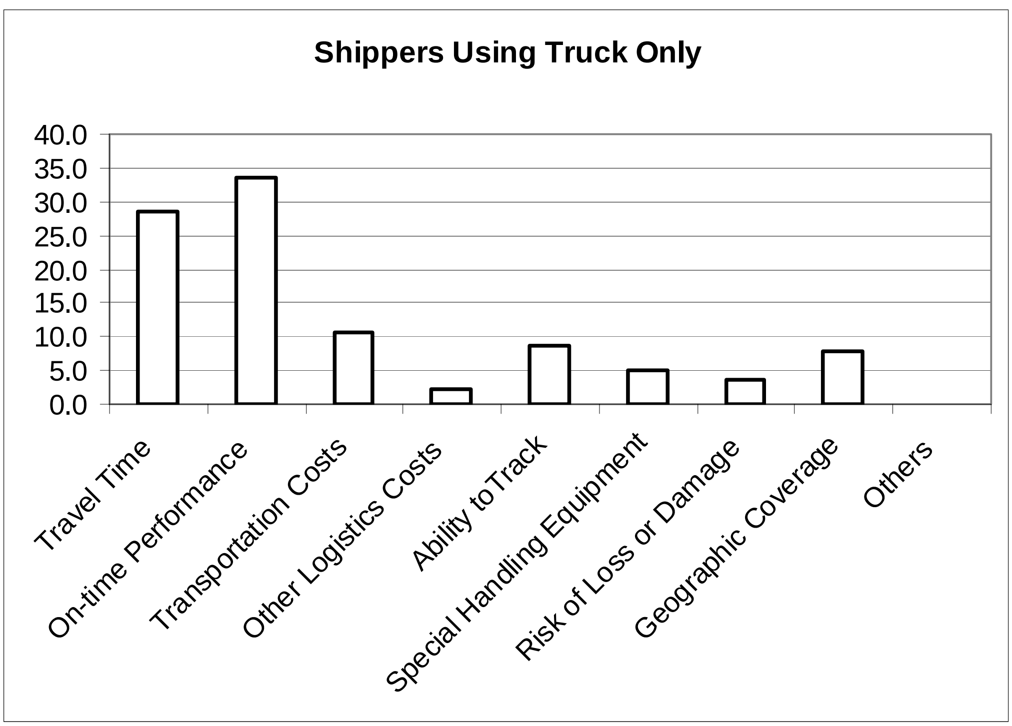


Figure 5.2: Average Relative Weights of Attributes for Shippers Using only Truck

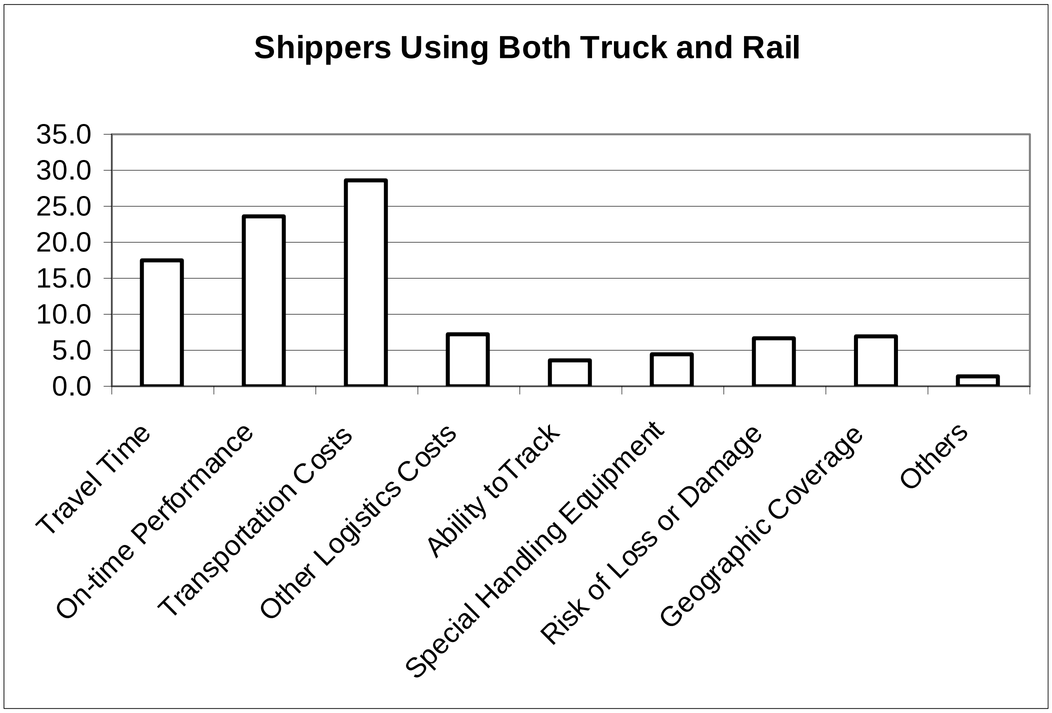


Figure 5.3: Average Relative Weights of Attributes for Shippers Using Truck and Rail

5.5 Relative Preferences by Commodity Type

The average relative weights given in Table 5.2 differ considerably for different commodity groups. Figure 5.4 provides the relative preferences among the attributes for the three different commodities STCC 2013, STCC 2631 and STCC 3711. Figures 5.5 and 5.6 show the relative preferences for the three commodity types for shippers that use truck only and for shippers that use both truck and rail respectively. For Fiber, Paper or Pulp Board manufacturers (STCC 2631) travel time, on-time performance and total logistics cost are the most important attributes accounting for about 70 percent of the total weight in determining the choice of mode. For Motor Vehicle manufacturers (STCC 3711) on-time performance and total logistics costs are the important factors accounting for about 50 percent of the total weight. For Meat Product manufacturers (STCC 2013) on-time performance and travel time are the most important factors account for about 75 percent of the total weight.

This section provides some plausible explanations for the above observations. Motor Vehicles are relatively high priced products. However, due to the global nature of their supply chains their total cycle times are relatively longer and travel time is not the most important attribute affecting the choice of mode. However, due to the sophisticated nature of their supply chain management practices like accurate sales forecasting techniques, on-time performance is important for motor vehicles. Even though meat products are relatively low priced products; travel time and on-time delivery are very important attributes for food products because of their perishable nature. Therefore meat product manufacturers do not use rail for shipping.

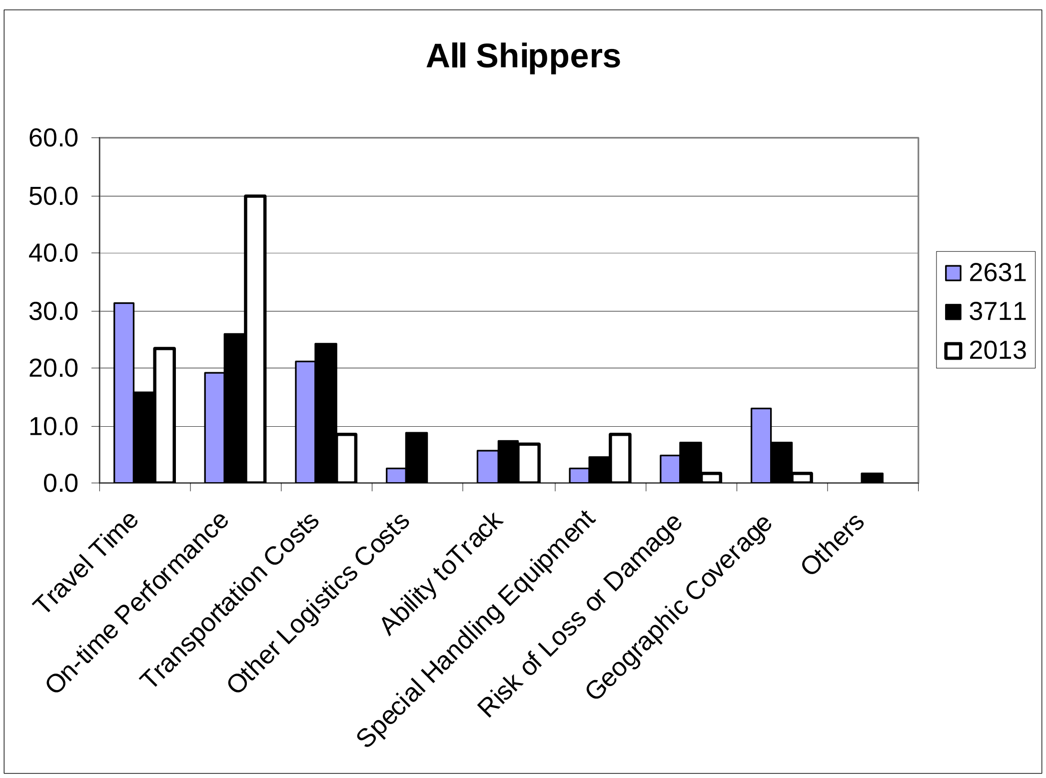


Figure 5.4: Commodity wise Average Relative Weights of Attributes for All the Shippers

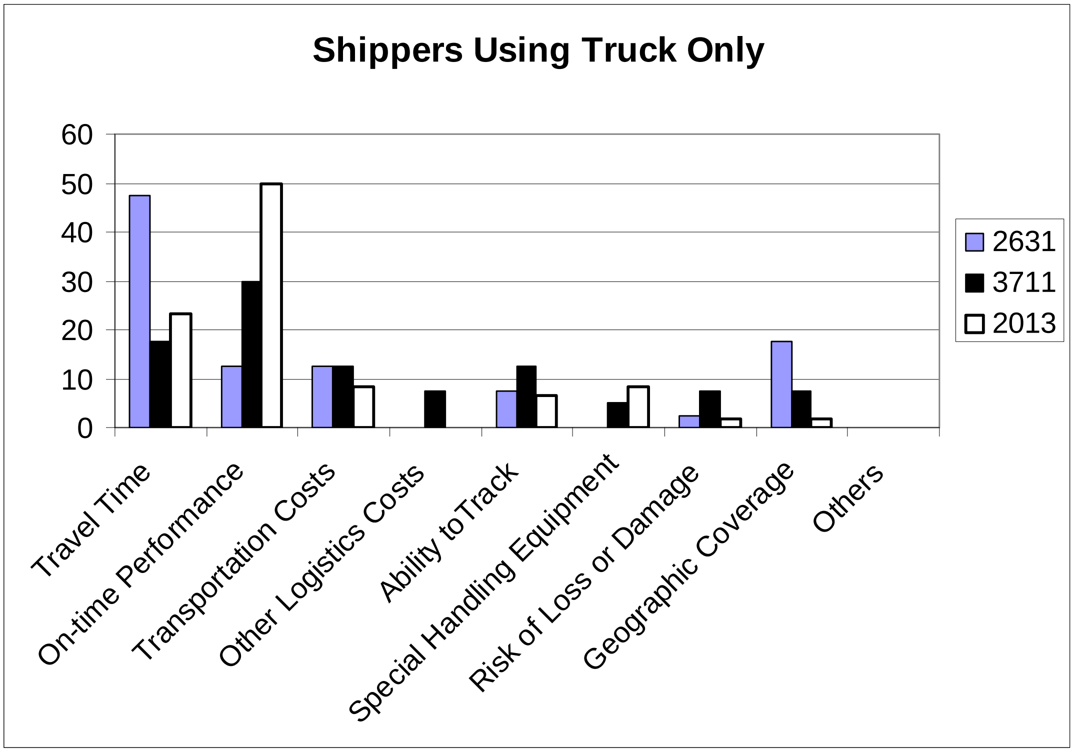


Figure 5.5: Commodity wise Relative Weights of Attributes for Shippers using only Truck

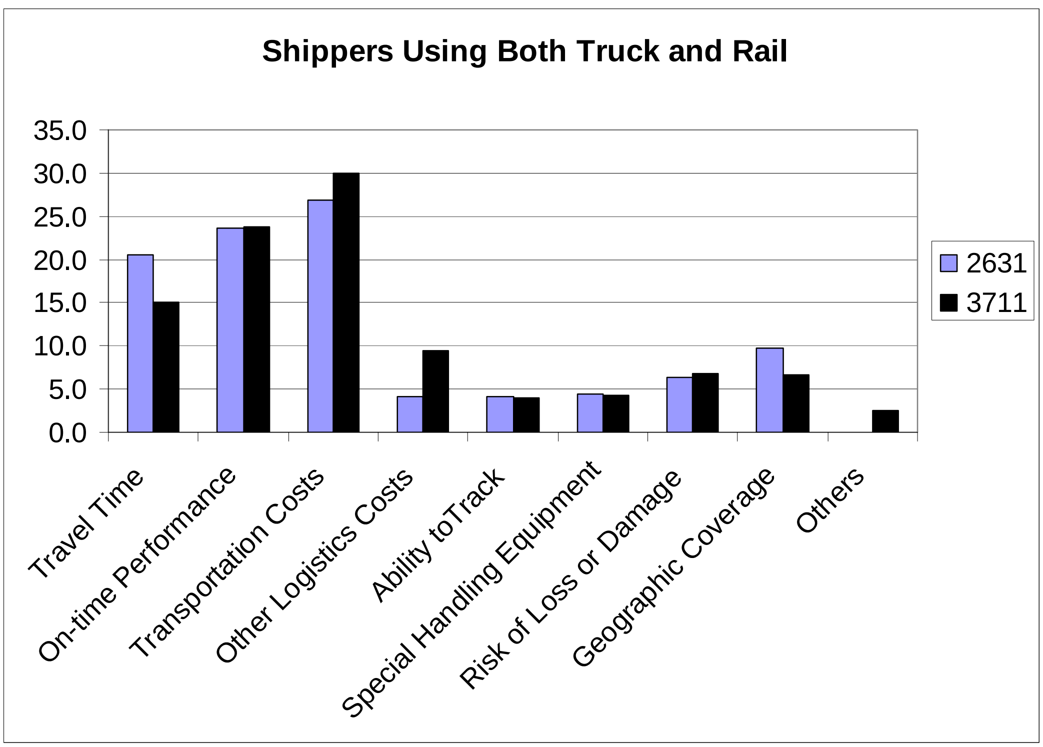


Figure 5.6: Commodity wise Relative Weights of Attributes for Shippers Using Both Truck and Rail

5.6 Performance of Truck versus Rail

The travel times, percentage of shipments on-time, transportation cost as a percentage of shipment value and other logistics costs as a percentage of shipment value were also obtained from the survey of shippers. These values were obtained for distances of 200 miles, 500 miles and 1000 miles for truck and rail. The travel times and percentage of shipments on-time were consistent across shippers from all the three commodities. Hence, the median values of travel time and travel time reliability were presented as the estimated travel time, on-time performance values for 200, 500 and 1000 miles.

Table 5.4: Comparison of Travel Time and On-Time Performance for Truck and Rail

	All Shippers					
Factor	200 miles		500 miles		1000 miles	
	Truck	Rail	Truck	Rail	Truck	Rail
Travel Time (in days)	0.5	2.8	1.0	4.0	2.0	6.0
On-time performance	99.0 %	80.0 %	98.0 %	70.0%	96.5%	65.0 %

The values obtained for transportation costs and other logistics costs differed significantly among the commodity groups. On comparison with the other responses within each commodity group, one of the responses for transportation costs and other logistics costs was identified as a major outlier as its value exceeded the mean of the other responses by more than three times. Hence the average value of each group after excluding a major outlier is presented in Table 5.5.

Table 5.5: Commodity Wise Comparison of Transportation and Other Logistics Costs for Truck and Rail

Factor	200 miles		500 miles		1000 miles	
	Truck	Rail	Truck	Rail	Truck	Rail
Fiber, Paper or Pulp Board (STCC 2631) Shippers						
Transportation costs (%)	11.0	7.5	13.5	9.5	17.0	12.0
Other logistics costs (%)	2.8	2.2	3.2	2.5	3.6	3.2
Motor Vehicle (STCC 3711) Shippers						
Transportation costs (%)	2.2	1.8	3.6	1.9	4.4	2.0
Other logistics costs (%)	1.1	1.1	1.1	1.1	1.3	1.1
Meat Product (STCC 2013) Shippers						
Transportation costs (%)	4.0	3.0	5.0	3.0	6.0	4.0
Other logistics costs (%)	1.0	2.0	1.0	2.0	1.0	2.0

The performance of truck is much better in terms of travel time and on-time service. However, rail performs marginally better in terms of transportation and other logistics costs.

5.7 Summary

The analytical method used in this chapter helped in understanding the relative preferences among different attributes that influence the choice of mode for different commodity groups. The shippers of Meat Products indicated that they do not use rail because rail does not provide refrigeration facilities. Hence, rail is not a feasible mode for perishable products even if the travel times and reliability were competitive. On-time performance and total logistics cost are the most important attributes that determine the choice of mode for Motor Vehicle manufacturers. In my opinion on-time performance is an attribute for which the performance of rail can be improved; in which case a significant number of Motor Vehicle shipments that are currently shipped by truck can be diverted towards rail. Travel time and total logistics are the most important attributes for Fiber, Paper or Pulp Board manufacturers. Improving the performance of rail for travel time or logistics cost is a difficult proposition unless expensive infrastructure improvements are undertaken. Hence, further diversion of the Fiber, Paper or Pulp Board shipments is difficult.

The analytical method presented in this chapter can be used in identifying the potential commodities for modal diversion and the improvements in transportation service required for the diversion. For example, if a new intermodal facility is being planned, then all the important commodity groups that move through the region need to be identified first. Then a survey similar to the one presented in this chapter can be sent out to a representative sample of shippers from each commodity group. Such a survey would be useful in identifying the factors that are most important for modal diversion for the important commodities moving through the region. Then the proposed intermodal facility should focus on improving the performance of rail and truck for these important factors.

However, in order to quantify how the performances of rail and truck on these important factors translate into the actual number of shipments used by each mode a more rigorous discrete choice modeling approach is required. The method presented in this chapter requires a separate analysis for each commodity group; however a discrete choice model can be used for all the commodity groups. Hence, discrete choice models are developed in the next chapter using the data available from the TRANSEARCH database and some of the data collected from the survey described in this chapter.

Chapter 6

Empirical Choice Modeling

6.1 Need for Empirical Choice Modeling

The development of empirical discrete choice models for modeling the choice of transportation mode has been an active area in transportation research over the past four decades. The use of discrete choice models is popular because of their high accuracy and sensitivity to policy measures. However, they are more data intensive as compared to the analytical method. Discrete choice models are useful for transportation planners for two important applications. The first application is in obtaining a Modal Split in the four step planning process for travel demand forecasting. The second application of discrete choice models is in policy analysis. They can be used as a tool in analyzing policy measures like studying the potential impacts of imposing tolls and calculating the potential benefits due to proposed improvements in transportation infrastructure. These policy measures can be used to affect modal shifts in order to improve the overall efficiency of the transportation system.

The discrete choice models that have been developed so far have been logit models for the most part. However, depending on the nature of data available and the primary purpose of developing the model some other classification models might be more appropriate. The use of less common models like Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Classification Trees for mode choice modeling is studied in this chapter. These models were selected because of their successful application to classification problems in fields like Medicine and Business.

6.2 Training Data Set and Test Data Set

Two separate data sets were prepared in this study to compare and understand the predictive ability of different modeling methods. The first data set, referred as the training data set from here on, is used for calibration of the models and the second data set, referred as the test data set from here on, is used for testing the accuracy of the models. The training data set is prepared based on the outbound shipment data from TRANSEARCH database for Arlington County and the test set is prepared based on outbound shipment data from TRANSEARCH database for King William County. County level outbound data is selected because the origin of the shipments would be known with a reasonable accuracy. This approach towards testing the model, allows us to test the performance of the model on an independent data set and it also looks at the transferability of the model across different geographic regions.

6.3 Preparation of Data Sets

Only the preparation of the training data set is described here as the training data set and test data set were prepared exactly in the same manner. TRANSEARCH database provides information on county to county annual flows for all the commodities at a four digit STCC commodity code level. This database provides total annual flows in and out of Virginia as well as within Virginia. The flows are provided separately for truck and rail shipments. The data pertaining to the outbound flows from Arlington County was queried from the TRANSEARCH database and was created as a separate data set.

This data set consists of the origin county (Arlington County), destination (a county, city, state or a BEA region), the four digit STCC commodity code, commodity flows by truck and rail. Only flows that have a county, city or state as a destination have been used. The destination states are approximately represented by the city closest to the state's centroid for calculating the Origin-Destination distances. The flows with a BEA region as a destination were excluded because these regions are too large in size to be approximated by a centroid. The distances were calculated for all the O-D pairs originating from the Arlington County. The TRANSEARCH database also provides the Value per Ton for all the four digit STCC commodity codes. The values corresponding to all the commodities in the Arlington data set were linked to it. At this stage, the training data set consisted of three variables that potentially affect the choice of mode, namely: shipment distance, total annual flow and commodity value. Though not an accurate measure, total annual flow can be considered as a surrogate measure for the size of the shipment. These three variables were used in the development of Model I using various methods that are described in the following sections.

An attempt has been made to combine information related to some alternative specific variables from the survey of shippers with the data from TRANSEARCH database. The survey obtained data related to travel time and reliability (percentage of shipments on-time) from the shippers for distances of 200, 500 and 1000 miles for Truck and Rail. Using a simple linear regression model with distance as the explanatory variable, the travel times and reliability values were estimated for the corresponding distances of Rail and Truck for all the O-D pairs in the Arlington data set. Similarly, the value of total logistics cost was estimated based on the shipment distance and commodity value for each of the O-D pairs in the Arlington data set. Now, the data set consists of three more variables namely: travel times for truck and rail, reliability estimates for truck and rail and total logistics costs for truck and rail. The following example illustrates how the travel time, reliability and total logistics cost were estimated for a typical observation.

Example: If an observation represents an annual flow of 6 tons for commodity STCC 2771 (Newspapers) between Arlington and Galaxy counties; the distance between the origin and destination is 320 miles and the value of the commodity is 3206 dollars/ton. Travel times and reliabilities for 320 miles were regressed on distance based on the data obtained from the survey for 200, 500 and 1000 miles. The travel time estimates for truck and rail are 0.88 days and 3.33 days respectively; reliability estimates are 97.3 % and 72.3 % respectively. Total logistics costs were regressed on commodity value and distance based on data obtained from the survey for three commodity values and distances of 200, 500 and 1000 miles. Total logistics costs were estimated as 10.44 % and 7.50 % of the shipment value for truck and rail respectively. Based on these estimates the actual total logistics costs are estimated as $ 33,471 and $ 24,045 respectively.

Model I developed using data only from the TRANSEARCH database is further improved by combining the data from the survey of shippers to develop Model II. Both Models I and II were used for all the four methods described in the following sections. The Arlington data set consisted of 850 observations and King William data set consisted of 859 observations. At this stage, the commodities were classified as perishables and non-perishables. The perishable commodities were excluded from both the data sets since all perishable commodities were being shipped by truck and they did not have a choice of mode. This resulted in a training set (Arlington data set) of 681 observations and a test set (King William data set) of 830 observations.

6.4 Development of a Binary Logit Model for Choice of Mode:

Binary logit models and binary probit models are two popular forms of binary discrete choice models. The use of logit models is very popular because logit models provide a convenient closed form solution to probabilities of choice. Though computationally straightforward, the logit models can be applied only when a property called as independence of irrelevant alternatives (IIA) is satisfied. While modeling the choice of more than two modes, sometimes IIA does not hold and logit models need to be used cautiously. However, in case of binary mode choice modeling, there are no complications associated with IIA and hence a logit model is used in this study.

A binary logit model consists of two utility functions that represent the total utility provided by each mode to the shipper. The utility functions, which are assumed to be linear in parameters, are presented below [34]:

$$U_{in} = \beta_1 x_{1in} + \beta_2 x_{2in} + \cdots + \beta_k x_{kin} + \varepsilon_i$$

$$U_{jn} = \beta_1 x_{1jn} + \beta_2 x_{2jn} + \cdots + \beta_k x_{kjn} + \varepsilon_j$$

Here, 'n' denotes the observation, 'i' and 'j' represent the two modes being considered, 'x' represents the variables identified above and 'β' represents the co-efficients of the parameters in the utility function. The logit model assumes that the error term 'ε' follows a logistic distribution. The probability that the mode 'i' is selected is given by the following expression:

$$P_n(i) = P(U_{in} > U_{jn})$$

$$P_n(i) = P_r(\varepsilon_n \le V_{in} - V_{jn}) = \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{in}}}$$

Now the calibration of the model involves obtaining the values of the co-efficients (β) in the utility functions. This has been done using the statistical software 'R'[2].

In binary mode choice models, the utility of one of the modes can be arbitrarily assigned zero because the probability of choosing a mode depends only the difference between the utilities of the two modes. In the calibration of models for the present study, the utility of rail is assigned zero. The differences between the values of alternative specific variables are used instead of the actual variable in the model calibration.

A preliminary model (Model I) has been developed using the variables obtained from the TRANSEARCH database: distance, value of the commodity and total tonnage. The model is calibrated using 'R' and the parameter estimates are shown in Table 6.1. The 'R' Codes used for all the models are provided in Appendix C.

Table 6.1: Logit Model Parameter Estimates for Model I

Variable	Estimate	Std. Error	Z Value	Pr(>|Z|)
(Intercept)	3.929	0.5454	7.203	5.91E-13
Distance	-0.00317	0.000962	-3.296	0.00098
Value	0.001395	0.0005	2.794	0.00521
Total Tonnage	-0.00036	6.8E-05	-5.361	8.26E-08

Here 'Estimate' denotes the parameter estimate of the explanatory variable, 'Standard Error' denotes the Standard Deviation of the sampling distribution of the estimate, Z- Value denotes the standardized value of the estimate and it is obtained by dividing the value of the estimate by the standard error and Pr(>|Z|) denotes the probability of the parameter estimate being insignificant or the value of the parameter estimate becoming zero.

[2] 'R' is an open source statistical programming language available freely under GNU General Purpose License.

Utility functions for Model I:

U_{truck} = 3.929 + 0.0013195 * (value) –0.00317*(distance) –0.00036*(total_tonnage)

U_{rail}= 0

The prediction accuracies of the model on the training set and test set are shown in Tables 6.2 and 6.3.

Table 6.2: Accuracy of Logit Model I with a probability threshold of 0.50

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	664	654	10	785	773	12
Accuracy(%)	97.50	99.54	41.67	94.58	94.15	41.38

Table 6.3: Accuracy of Logit Model I with a probability threshold of 0.75

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	662	648	14	772	756	16
Accuracy(%)	97.21	98.63	58.33	93.01	92.08	55.17

This model has been further improved by incorporating additional data from the survey of shippers. The variables to be considered are: distance, total tonnage, commodity value, difference in travel time, difference in reliability and difference in total logistics costs. The correlation matrix involving all the above variables is shown in Table 6.4.

Table 6.4: Correlation Matrix for All the Explanatory Variables

	Distance	Value	Total Tonnage	Diff. TT	Diff. Rel.	Diff. Cost.	Choice
Distance	1.0000	0.0426	-0.0231	-1.0000	0.8094	-0.0262	-0.0710
Value	0.0426	1.0000	-0.0811	-0.0426	-0.0627	-0.0243	0.0671
Total Tonnage	-0.0231	-0.0811	1.0000	0.0231	0.0235	0.4145	-0.5366
Diff. TT	-1.0000	-0.0426	0.0231	1.0000	-0.8094	0.0262	0.0710
Diff. Rel.	0.8094	-0.0627	0.0235	-0.8094	1.0000	-0.0458	-0.1334
Diff Cost.	-0.0262	-0.0243	0.4145	0.0262	-0.0458	1.0000	-0.0656
Choice	-0.0710	0.0671	-0.5366	0.0710	-0.1334	-0.0656	1.0000

The variables distance, difference in travel time and difference in reliability exhibit a very high correlation because the values of travel time and reliability were estimated based on the distance. These variables should not be simultaneously used in the model. Therefore, only difference in travel time is used in Model II instead of the distance. Model II has been developed using the variables commodity value, total tonnage, difference in travel time and difference in total logistics cost. The parameters calibrated for the model are tabulated in Table 6.5.

Table 6.5: Logit Model Parameter Estimates for Model II

	Estimate	Std. Error	Z value	Pr(>|Z|)
(Intercept)	5.61E+00	1.06E+00	5.307	1.11E-07
Value	1.87E-03	9.37E-04	1.996	0.0459
Total Tonnage	-3.33E-04	7.76E-05	-4.293	1.76E-05
Diff. TT	-1.19E+00	3.64E-01	3.252	0.00115
Diff. Cost.	-1.71E-08	2.38E-08	-0.719	0.47189

Utility functions for Model II:

U_{truck} = 5.61 + 0.00187* (value) – 0.000333*(total_tonnage) –1.19 (travel_time_truck) – 1.71*(total_logistics_cost_truck)

U_{rail}= –1.19 (travel time for rail) – 1.71*(tota_logistics_cost_rail)

The prediction accuracy of the model on the training set and test set are shown in Tables 6.6 and 6.7.

Table 6.6: Accuracy of Logit Model II with a probability threshold of 0.50

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	663	654	9	784	773	11
Accuracy(%)	97.36	99.54	37.50	94.46	94.15	37.93

Table 6.7: Accuracy of Logit Model II with a probability threshold of 0.75

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	662	648	14	772	756	16
Accuracy(%)	97.21	98.63	58.33	93.01	92.08	55.17

The prediction accuracy of Model I was marginally better than the prediction accuracy of Model II with a probability threshold of 0.50. The prediction accuracy of both the models is the same with a probability threshold of 0.75. Model I is used to draw inferences about the choice of mode because is it based on more reliable data. The distances at which shippers begin to prefer rail over truck as a function of the product value per ton and annual tonnage between an O-D pair are shown in Table 6.8. This table shows that rail is generally used for shipments whose annual tonnage between an O-D pair is greater than 10,000 tons and whose value is less than 3,200 dollars per ton.

Table 6.8: Distances at Which Shippers Begin to Prefer Rail for Various Product Values and Annual Tonnages

Value \ Tons	10	50	100	250	500	1000	2000	5000	10000	15000	20000	30000	40000
50	1259	1255	1249	1232	1203	1147	1033	692	125	Min.	Min.	Min.	Min.
100	1280	1275	1270	1253	1224	1167	1054	713	145	Min.	Min.	Min.	Min.
150	1301	1296	1291	1273	1245	1188	1075	734	166	Min.	Min.	Min.	Min.
200	1322	1317	1311	1294	1266	1209	1096	755	187	Min.	Min.	Min.	Min.
400	1405	1400	1395	1378	1349	1292	1179	838	270	Min.	Min.	Min.	Min.
800	1571	1567	1561	1544	1516	1459	1345	1005	437	Min.	Min.	Min.	Min.
1600	1904	1900	1894	1877	1849	1792	1678	1338	770	202	Min.	Min.	Min.
3200	2570	2566	2560	2543	2515	2458	2344	2004	1436	868	300	Min.	Min.
4800	3236	3232	3226	3209	3181	3124	3010	2670	2102	1534	966	Min.	Min.
6400	3902	3898	3892	3875	3847	3790	3676	3336	2768	2200	1632	496	Min.
8000	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	3434	2866	2298	1162	Min.
10000	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	3698	3131	1995	859
20000	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.
40000	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.

Notes:

1) In the above table, N.P. refers to Not Preferred, i.e. Rail is not preferred for these shipments for any distance less than 4,000 miles.

2) In the above table, Min. refers to Minimum, i.e. Rail is preferred for these shipments at any minimum distance for which the rail operations are feasible.

6.5 Mode Choice Modeling using Linear Discriminant Analysis (LDA)

In case of binary choice modeling Linear Discriminant Analysis attempts to find a hyperplane that separates the p-dimensional space into two halves [35]. Here the p-dimensions represent each of the explanatory variables that affect the choice of mode. The points that lie on one side of the place represent the truck mode and the points that lie on the opposite side represent the rail mode. LDA is a special case of the general discriminant problem that assumes that covariance matrices of all the classes are equal.

If we represent the observations for the two choices as classes 'k' and 'l', the linear discriminant function for class 'k' can be represented by:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k{}^T \Sigma^{-1} \mu_k + \log \pi_k$$

The decision boundary between the classes 'k' and 'l' is described by:

G(x)=argmax$_k$ $\delta_k(x)$. This can be denoted by the following linear equation:

$$\log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^T \Sigma^{-1}(\mu_k - \mu_l) + x^T \Sigma^{-1}(\mu_k - \mu_l) = 0$$

If the value of the above expression is greater than zero, the observation is classified as truck and if it is less than zero the observation is classified as rail. Here x represents an observation written as a vector of p explanatory variables, π_k and π_l represent the proportion of observations in classes k and l, μ_k and μ_l represents the class mean vectors and Σ represents the common covariance matrix for all classes.

The above parameters can be estimated as:

$\hat{\pi}_k = N_k / N$, where N_k is the number of class-k observations

$\hat{\mu}_k = \sum_{g_i = k} x_i / N_k$

$\hat{\Sigma} = \Sigma_{k=1}^{K} \Sigma_{g_i = k} (x_i - \mu_k)(x_i - \mu_k)^T / (N - K)$

For a detailed description on LDA, please refer to References [36, 37].

The above parameters are estimated using R and the co-efficients of the explanatory variables are tabulated in Tables 6.9 and 6.12 for Models I and II.

6.5.1 Model I

Table 6.9: Co-efficients of Linear Discriminants for Model I

	LD1
Value	3.41E-06
Total Tonnage	-3.92E-04
Distance	-6.62E-04

The prediction accuracies of Model I on the training set and test set are shown in Tables 6.10 and 6.11.

Table 6.10: Accuracy of LDA Model I with default prior probabilities π_k and π_l

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	658	647	11	779	766	13
Accuracy(%)	96.62	98.48	45.83	93.86	93.30	44.83

Table 6.11: Accuracy of LDA Model I with probabilities $\pi_k = 0.25$ and $\pi_l = 0.75$

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	655	644	11	776	761	15
Accuracy(%)	96.18	98.02	45.83	93.49	92.69	51.72

6.5.2 Model II

Table 6.12: Co-efficients of Linear Discriminants for Model II

	LD1
Value	3.09E-06
Total Tonnage	-4.38E-04
Diff. TT	-2.29E-01
Cost. Diff	2.80E-08

The prediction accuracies of Model II on the training set and test set are shown in Tables 6.13 and 6.14.

Table 6.13: Accuracy of LDA Model II with default prior probabilities π_k and π_l

	Training Set			**Test Set**		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	660	649	11	777	764	13
Accuracy(%)	96.92	98.78	45.83	93.61	93.06	44.83

Table 6.14: Accuracy of LDA Model II with probabilities $\pi_k = 0.25$ and $\pi_l = 0.75$

	Training Set			**Test Set**		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	658	647	11	775	760	15
Accuracy(%)	96.62	98.48	45.83	93.37	92.57	51.72

The prediction accuracy of models I and II for both prior probabilities is nearly the same.

6.6 Mode Choice Modeling using Quadratic Discriminant Analysis (QDA)

QDA uses a quadratic discriminant surface to separate the p-dimension space into two halves. QDA arises when the assumption of the equality of covariance matrices among all the classes is relaxed. The following equation represents a quadratic discriminant function for class 'k':

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \log \pi_k$$

The decision boundary between the classes k and l is represented by the quadratic equation: $\{ x : \delta_k(x) = \delta_l(x) \}$

The prediction accuracies of QDA on the training set and test set are shown in Tables 6.15 and 6.16. Only the default prior probabilities were shown below as a deviation from default prior probabilities is significantly decreasing the accuracy percentage.

6.6.1 Model I:

Using variables: distance, commodity value and total tonnage

Table 6.15: Accuracy of QDA Model I with default prior probabilities π_k and π_l

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	663	644	19	758	738	20
Accuracy(%)	97.36	98.02	79.17	91.33	89.89	68.97

6.6.2 Model II

Using variables: commodity value, total tonnage, difference in travel time and difference in total logistics cost

Table 6.16: Accuracy of QDA Model II with default prior probabilities π_k and π_l

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	659	637	22	742	718	24
Accuracy(%)	96.77	96.96	91.67	89.40	87.45	82.76

The overall prediction accuracy of models I and II is nearly the same but model II performs better than model I for observations with rail as their choice.

6.7 Mode Choice Modeling using Tree Based Methods

Classification Trees are simple but powerful tools used in modeling choices. A tree consists of a series of nodes which hierarchically classify the observations into groups. At each node, the observations are split into two groups based on a threshold value of a particular explanatory variable. These groups are hierarchically further split into groups; two groups at a time based on threshold values of other explanatory variables. At the final set of nodes referred to as the terminal nodes; the observations are classified as belonging to one of the choices. The calibration of a tree involves developing a full tree that gives the best possible classification on the training data set and pruning[3] the tree to a reasonable level to avoid over fitting[4]. Tree pruning is analogous to eliminating some of the insignificantly contributing variables in regression modeling. Trees can be pruned using statistical procedures like Cross-validation, Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). For a detailed description of Tree Classification, please refer to Reference [37].

Initially a fully grown tree using the variables total tonnage (flow), difference in travel time, commodity value and difference in total logistics costs is developed using

---

[3] Pruning refers to the process of reducing the number of nodes in a Classification Tree to improve the performance of the model outside the training data set.

[4] The presence of too many nodes leads to a model "over fit", i.e. the model excessively "fits" the training data set and performs very well on the training dataset. However, the model loses its generality and performs badly on the test set which is not a desirable quality for the model.

'R'. The figure 6.1 shows a fully grown tree. A detailed tree classification output is provided in Appendix D.

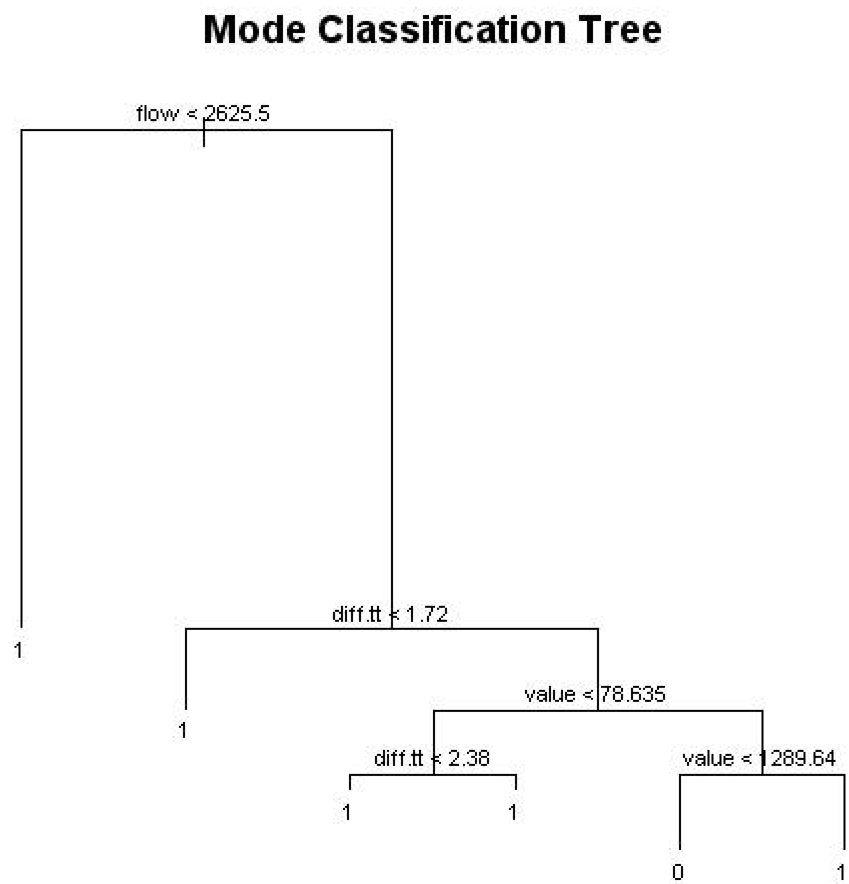


Figure 6.1: A Fully Grown Tree

6.7.1 Tree pruning using cross-validation

The mis-classification rate versus number of nodes plot has been drawn and is shown in Figure 6.2. Based on this plot, it has been decided that a five node tree will be the most suitable tree classification model for this study.

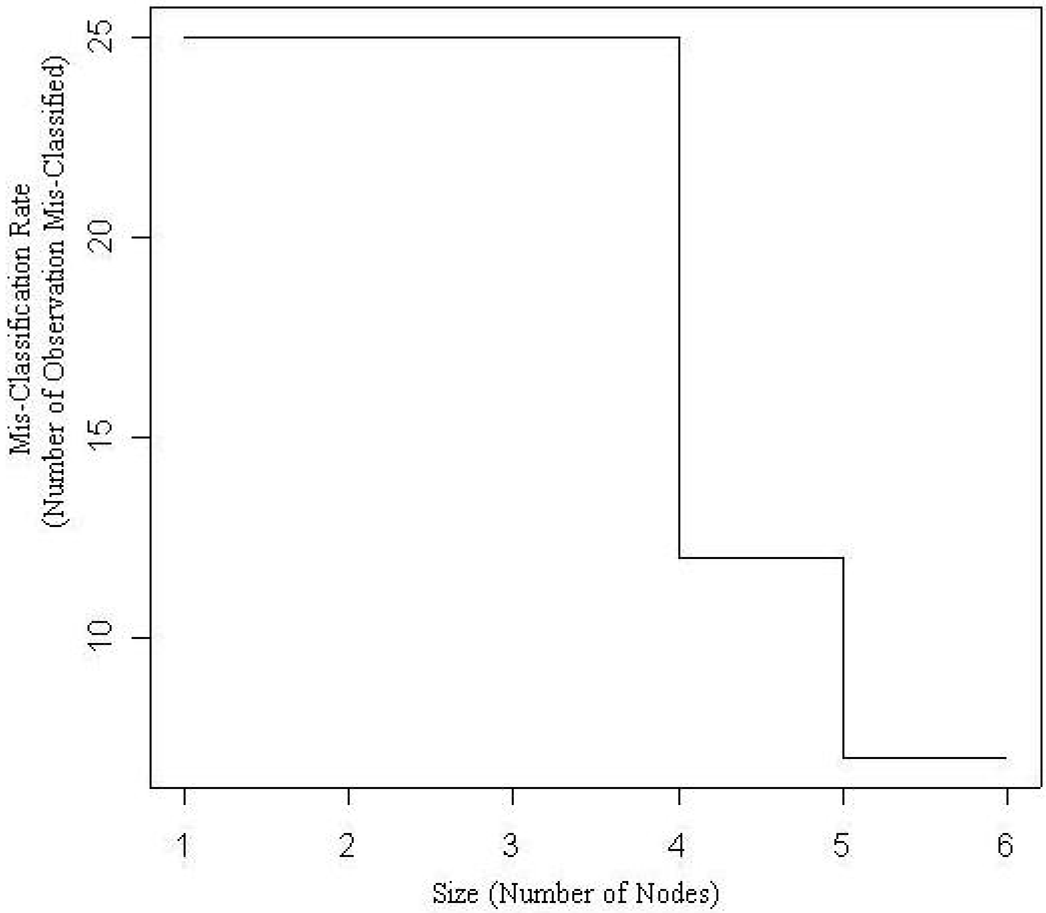


Figure 6.2: Mis-Classification Rate versus Tree Size

The five-node tree shown in the following page is used as the final tree classification model. The prediction accuracy of this model on the training and test data sets is shown Table 6.17.

Table 6.17: Prediction Accuracy of a Five Node Tree

	Training Set			Test Set		
	Total	Truck	Rail	Total	Truck	Rail
Actual	681	657	24	830	821	29
Correct	678	657	21	791	761	22
Accuracy(%)	99.56	100.00	87.50	95.30	92.69	75.86

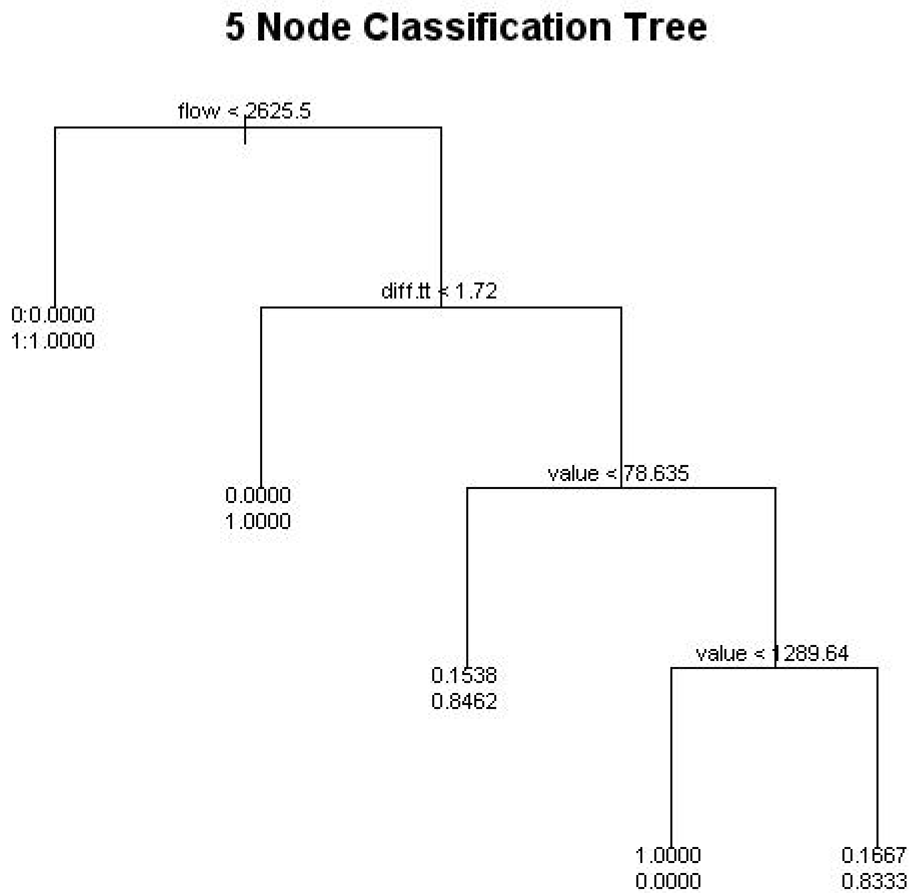


Figure 6.3: A Five Node Classification Tree

The resultant trees using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are shown in Appendix D.

6.8 Summary

Mode choice modeling using four different binary choice analysis methods was done in this chapter. The performance of Models I and II was nearly the same for Logit Model, LDA and QDA. The use of model II is recommended because it accounts for two important variables: travel time and total logistics cost. The prediction accuracy of all the four methods on the training set and the test set for Model II is compared in Tables 6.18 and 6.19.

Table 6.18: Comparison of Prediction Accuracy of all Four Methods on Training Set

	All Observations (Total)				Only Rail			
	Logit	LDA	QDA	Trees	Logit	LDA	QDA	Trees
Actual	681	681	681	681	24	24	24	24
Correct	662	658	659	678	14	11	22	21
Accuracy(%)	97.21	96.62	96.77	99.56	58.33	45.83	91.67	87.50

Table 6.19: Comparison of Prediction Accuracy of all Four Methods on Test Set

	All Observations (Total)				Only Rail			
	Logit	LDA	QDA	Trees	Logit	LDA	QDA	Trees
Actual	830	830	830	830	29	29	29	29
Correct	772	775	742	791	16	15	24	22
Accuracy(%)	93.01	93.37	89.40	95.30	55.17	51.72	82.76	75.86

The overall prediction accuracy of all the four methods is relatively high. This is mainly because of the fact that most of the observations belong to truck mode. Hence, the prediction accuracy for observations with the choice of rail needs to be considered carefully while adopting a model. The overall prediction accuracy of all the four methods is very high (in excess of 95%) on the training set. Tree Classification and Logit models have shown the highest overall prediction accuracy on the test set. The overall prediction accuracy of LDA and QDA are also reasonably good on the test set. However, when we consider only those observations with rail as their mode, the prediction accuracy of Logit model and LDA is low. The prediction accuracy of QDA is the highest for rail and the prediction accuracy of Classification Trees is also reasonably high.

Chapter 7

Conclusions

7.1 Summary

A two step modeling methodology that attempts to overcome some of the deficiencies in the previous freight planning modeling efforts has been illustrated. The first step of this methodology is equivalent to the trip generation and distribution steps of the 4-step planning process. This substitution was necessary because the process of trip generation and trip distribution as used for modeling passenger O-D flows is not directly applicable for modeling freight flows. This is because of the fact that the amount of freight being generated or attracted into a region cannot be usually explained by socio-economic variables of a region like the population of the region, number of employees etc.; and the use of regression models for freight trip generations would not be appropriate. Only, the trip attractions of consumer related goods can be modeled using the socio-economic variables. Hence, an alternative method of obtaining the O-D flows by tracing the supply chains of major business units in a region is suggested. A database like InfoUSA which provides a commodity wise listing of businesses in an area can be used to identify the important freight generating businesses in an area. This could be used in combination with factors like market share for the firm, size of the firm and total sales volume etc. to obtain the O-D flows. This step has been illustrated using a case study of Volvo's truck manufacturing plant in Virginia's Pulaski County. Publicly available information related to Volvo's supply chain and annual sales volumes is used in this case study. The illustration also helped in identifying certain errors in the TRANSEARCH database such as incorrect Origin-Destination pairs for STCC 3711 (Motor Vehicles).

The second step involves modeling the choice of mode for freight shipments. The logistical needs and constraints of a shipper determine the choice of mode. Therefore, a model that accounts for the logistical variables would be appropriate for modeling the choice of mode. A list of supply chain variables that have the potential to influence the choice of mode is identified. A survey of shippers was conducted to analyze the relative importance of some of the important supply chain variables on the choice of mode. Shippers of three different commodities: Motor Vehicles (STCC code 3711), Fiber, Paper or Pulp Board (STCC code 2631) and Meat Products (STCC code 2013) were surveyed and the differences in their preferences analyzed. An analytical method is useful understanding the preferences of various shippers, however, it not useful in converting these preferences into a numerical modal split among the freight shipments. Hence, the need for an empirical choice model is recognized and an attempt has been made to develop a discrete choice model.

A common problem that is usually reported in modeling the choice of mode is the lack of availability of reliable disaggregate data. A discrete choice model has been developed using aggregate data from TRANSEARCH database supplemented with non-sensitive information from a survey of shippers. The mode choice model was developed using four different classification techniques, namely: Binary Logit Model, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Tree Classification. The performance of four different discrete choice modeling techniques is compared on a training data set and a test data set.

Among the four techniques, LDA or QDA usually give good prediction accuracy. However, they do not give good interpretability for the variables; i.e. they do not provide the relative importance of the variables in the selection of mode. Tree classification is the simplest among the four methods and hence they are easiest to understand. However, it does not work well on certain types of data sets and does not provide the relative importance of the variables in determining the choice of mode. Logit models provide the best interpretability among the four methods because their co-efficients are useful in understanding the effect of each of the variables on the choice of mode. However, their predictive accuracy may be sometimes low if the distribution of the error terms does not follow the logistic distribution.

The use of Quadratic Discriminant Analysis (QDA) or Classification Trees is recommended for mode choice modeling if it is being done as a part of the four step planning process for obtaining modal splits. This is because in case of the four step planning process modal split accuracy is important and not the underlying reasoning behind the splits. However, it is advisable to compare the performance of the four methods on a test data set before adopting one of the methods because the accuracy of the methods also depends on the nature of the underlying data. The use of Logit models is recommended for mode choice modeling if it is being done for developing policy measures because it is based on economic theory and it can be used to evaluate the potential impacts of proposed policy measures.

Though the empirical mode choice models developed in this study are able to obtain modal splits with a good accuracy, they are still short of precisely accounting for the contribution of all the important factors in the mode choice decision process. When empirical choice models are able to account for the contribution of each of the factors they would be very useful for policy analysis. For developing empirical models that precisely account for all the important factors, data regarding additional factors needs to be obtained using a more elaborate questionnaire.

The major limitation of this kind of freight planning methodology is that it is data intensive and the collection of the required data can become a tedious and expensive process.

7.2 Conclusions

The following are the important conclusions from this study:

- The commodity flows presented in the TRANSEARCH database at a four digit STCC level are not always accurate.
- It is not always possible to protect the confidentiality of the data when commodity flows are published at a county level for four digit STCC codes.
- Travel time, on-time performance and transportation costs are the most important factors affecting the choice of mode accounting for about 70 percent of the total weight among all the factors.
- Quadratic Discriminant Analysis and Classification Trees provide the most accurate modal split among the four empirical choice models.
- Logit Models provide the most interpretable results among the four empirical choice models.
- Rail is usually the preferred mode for shipments whose value is less than 3200 dollars per ton and annual tonnage is greater than 10,000 tons.

- This methodology can be applied to statewide freight commodity flow forecasting either as a standalone methodology or in conjunction with the previous studies [3, 4]. It were used in conjunction with the previous studies it can be used to improve the commodity flows obtained from these studies.

7.3 Applications for Statewide Freight Transportation Planning

The methodology presented in this study can be used for statewide freight transportation planning. The key commodities moving in and out of Virginia were identified and trip generation equations were developed for these commodities in a previous study by Brogan [3]. Trip distribution equations were developed in another study by Mao [4]. These steps are useful in obtaining the commodity wise O-D flows for the shipments originating and terminating in Virginia. These steps were performed as a part of the system inventory step of the Statewide Intermodal Freight Transportation Planning Methodology. The method of obtaining the O-D flows, as illustrated in Chapter 4, can be used to improve upon the accuracy of the O-D flows obtained through the previous studies. The empirical mode choice models, developed in Chapter 6, are useful in obtaining the modal splits when O-D flows are obtained at a four digit STCC commodity level. These commodity flows for each mode will be useful in completing the "System Inventory" step of the Statewide Intermodal Freight Transportation Methodology. Apart from the System Inventory Step, the mode choice analysis performed using the Analytical method or by using the Logit Model will be useful for policy analysis like developing modal diversion measures. This analysis is useful in the

"Development and Evaluation of Improvement Alternatives" Step of the Statewide Intermodal Freight Transportation Methodology.

7.4 Recommendations for Future Research

- The applicability of the proposed method of obtaining O-D flows by tracing the supply chains of firms needs to further examined at a larger geographic level like a state or a BEA region.
- The impact of additional explanatory variables like reliability of transportation time, transportation time as a fraction of the product cycle time on the choice of mode need to be further understood. The feasibility of inclusion of some of these variables in the Commodity Flow Survey (CFS) needs to be examined.
- The use of Delphi techniques in combination with revealed preference data is recommended in future research related to freight mode choice modeling as it can be used to overcome the problems associated with multi-collinearity and confidentiality of data.
- Accessing the micro data corresponding to the Commodity Flow Survey (CFS) from the Center from Economic Studies (CES) is recommended for future freight modeling efforts as this micro data contains reliable shipment level information.

References

1. Bureau of Transportation Statistics, U. S. Department of Transportation, *2002 Commodity Flow Survey*, December 2004.
2. Eatough, C.J., Brich, S.C., and Demetsky, M.J., *A Methodology for Statewide Intermodal Freight Transportation Planning*, VTRC 99-R 12, Virginia Transportation Research Council, December 1998.
3. Brogan, J.J., *Application of Statewide Intermodal Freight Transportation Methodology*, Master of Science Thesis, University of Virginia, 2001.
4. Mao, S., Calibration of the Gravity Model for Truck Freight Flow Distribution, Master of Science Thesis, University of Virginia, 2002.
5. http://www.bus.iastate.edu/mcrum/TRLOG%20462/Spring%202003/2, *Transportation Overview and Update*, Accessed on 10th October 2003.
6. http://ops.fhwa.dot.gov/freight/pp/bts.pdf, *Expenses per Mile for the Motor Carrier Industry: 1990 through 2000 and Forecasts through 2005*, Accessed on 10th October 2003.
7. Engel, C., *Competition Drives the Trucking Industry*, Monthly Labor Review, Bureau of Labor Statistics, April 1998.
8. Regan, A., Veras, J.H., Chow G., and Sonstegaard, M.H., *Freight Transportation Planning and Logistics*, Transportation in the New Millennium: State of the Art and Future Directions, Transportation Research Board, National Research Council, January 2000.

9. Roberts, P.O., *Supply Chain Management: New Directions for Developing Economies*, Accessed on April 10, 2005. http://www.worldbank.org/html/fpd/transport/ports/tr_facil.htm

10. Davis, H.W., and Drum, W.H., *Logistics Cost and Service 2004*, Proceedings of the Council of Logistics Management 2004 Annual Conference, Philadelphia, October 2004.

11. Blinick, N., *Global Inventory Management - A Strategic Imperative Creating Flexible*, *Lean Global Supply Chains*, Proceedings of the Council of Logistics Management 2004 Annual Conference, Philadelphia, October 2004.

12. Cook, J.A., Holcomb, M.C., Manrodt, K.B., and Ross, T.J., *Trends and Issues in Logistics and Transportation: Twelfth Annual Survey*, Proceedings of the Council of Logistics Management 2003 Annual Conference, Chicago, September 2003.

13. Florida Department of Transportation, *Urban Highway Freight Modeling Including Intermodal Connectors for Florida*, Tallahassee, Florida, 2002.

14. Middendorf, D.P., Jelavich, M., and Ellis, R.H., *Development and Application of* Statewide, *Multimodal Freight Forecasting Procedures for Florida*, Transportation Research Record, No. 889, 1982.

15. Eusebio, V., and Rindom, S., *Interstate Movements of Manufactured Goods in Kansas*, Kansas Department of Transportation, May 1991.

16. Transportation Research Board, *NCHRP Synthesis 230: Freight Transportation Practices in the Public Sector*, Washington D.C., 1996.

17. Transportation Research Board, *NCHRP Report 388: A Guidebook for Forecasting Freight Transportation Demand*, Washington D.C., 1997.

18. Winston, C., *The Demand for Freight Transportation: Models and Applications*, Transportation Research, Vol. 17A, No. 6, 1983.

19. Abdelwahab, W., and Sargious, M., *Modeling the Demand for Freight Transport*, Journal of Transport Economics and Policy, January 1992.

20. Interviews with Logistics Managers of Three Major Retailers, conducted by Tatineni, V.C., March & April 2004.

21. Sen, S., Prozzi, J., and Bhat, C.R., *The Delphi Technique: An Application To Freight Mode Choice Analysis*, Proceedings of 84th Transportation Research Board Meeting, Washington D.C., January 2005.

22. Danielis, R., Marcucci, E., and Rotaris, L., *Logistics Managers' Stated Preferences for Freight Service Attributes*, Transportation Research Vol. 41 E, May 2005.

23. Louviere, J.J., Hensher, D.A., and Swait, J.D., *Stated Choice Methods: Analysis and Application*, Cambridge University Press, 2000.

24. Boerkamps, J.H.K., Van Binsbergen, A.J., and Bovy, P.H.L., *Modeling Behavioral Aspects of Urban Freight Movement in Supply Chains*, Transportation Research Record 1725, TRB, National Research Council, Washington D.C., 2000.

25. Tavasszy, L.A., Smeenk, B., and Ruijgrok, C. J., *A DSS for Modeling Logistic Chains in Freight Transport Policy Analysis*, Proceedings of the Seventh International Special Conference of IFORS: "Information Systems in Logistics and Transportation", Gothenburg, Sweden, June 1997.

26. Volvo Trucks North America History,
http://www.volvo.com/trucks/na/en-us/about_us/history/,
Accessed on 02/07/2005.

27. Volvo Products and Services,
http://www.volvo.com/group/global/en-gb/productsandservices/ ,
Accessed on 02/07/2005.

28. 2000th Mack Truck Rolls Out of New River Valley Plant,
http://www.macktrucks.com/default.aspx?pageid=891,
Accessed on 02/07/2005.

29. ArvinMeritor-Volvo Sign Three Year Cam Brake Deal,
http://www.todaystrucking.com/displayarticle.cfm?ID=3720,
Accessed on 02/12/2005.

30. Dealers and Service Locations,
http://www.volvo.com/trucks/na/en-us/dealers/,
Accessed on 02/12/2005.

31. Logistics the Volvo Way,
http://www.automotivewebtv.com/moxie/logistics/147.shtml,
Accessed on 02/12/2005.

32. Volvo Trucks Annual Reports for the Years 1998, 1999, 2000, 2001, 2002 and 2003, http://www.volvo.com/group/global/en-gb/investors/financial_reports/
Accessed on 02/07/2005.

33. Gibson, B.J., and Wilson, J.W., *Saturn Corporation: Improving the Plant-Retailer Link in the Auto Industry Supply Chain*, A Case Study Available from the Council of Supply Chain Management Professionals (CSCMP) Online Resources.

http://www.cscmp.org/Downloads/CaseStudy/saturn96.pdf, Accessed on 10th May, 2005.

34. Ben-Akiva, M., and Lerman, S.R., *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge: MIT Press, 1985.
35. http://en.wikipedia.org/wiki/Linear_classifier, *Linear Classifier*, Wikipedia, Accessed on 27th May 2005.
36. http://www.statsoft.com/textbook/stathome.html, StatSoft Electronic Textbook, *Discriminant Analysis*, Accessed on 27th May 2005.
37. Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning*: *Data Mining, Inference and Prediction,* Springer Series in Statistics, 2001.

Appendix A: Preliminary Questionnaire

1) Please provide the following details pertaining to your mode choice decision process:

a) Who makes the mode choice decision?

- ☐ Shipper
- ☐ Receiver
- ☐ Joint decision by both shipper and receiver

b) What are the most important modal attributes? (Distribute 80 points among these attributes on the basis of their importance)

	Attribute	Score
i.	On time performance	
ii.	Transit time	
iii.	Price	
iv.	Ability to track the status of the shipment	
v.	Availability of special equipment	
vi.	Risk of loss/damage	
vii.	Geographic coverage	
viii.	Other (specify)	
	Total Points	**80**

c) Provide your perceived estimates of the following attributes for Rail and Truck for a typical shipment.

	Rail	Truck
Travel time		
Travel time reliability (% of on-time delivery)		
Transportation costs as a proportion of shipment value		
Other logistics costs as a proportion of shipment value*		

*Other logistics costs associated with the shipment include:

i) Order processing costs
ii) Product handling and storage costs
iii) Capital costs of goods in inventory and transit
iv) Stock out costs in case of late shipments

2) Please provide the details pertaining to five typical outbound shipments originating from your establishment during normal season and peak season in the tables provided. If your establishment uses both rail and truck modes for transportation, include shipments carried by both truck and rail.

Definition of a Shipment: A shipment is a single movement of goods, commodities, or products from an establishment to a single customer or to another establishment owned or operated by the same company as the originating establishment (e.g., a warehouse, distribution center, or retail or wholesale outlet). Full or partial truckloads are counted as a single shipment only if all commodities on the truck are destined for the same location. If a truck makes multiple deliveries on a route, each stop is counted as a separate shipment.

Explanation to various terms in the tables is provided below:

a) Choice of mode: Truck (T)/ Rail (R) /Both Tuck and Rail (T&R)

b) Shipment distance: The distance traveled by the shipment in miles

c) Shipment weight: The weight of the shipment in pounds

d) Shipment volume: The volume of shipment in cubic feet or as a fraction of a truckload

e) Frequency of shipment for the above destination

f) Value of the shipment (excluding transportation costs)

g) Transportation time of the shipment

h) Inventory storage time of the shipment within your establishment

i) Total product cycle time: The time elapsed between order placement and order delivery

j) Transportation costs per shipment

k) Other logistics costs associated with the shipment (If you do not have the absolute value of the logistics costs, please indicate them as a percentage of the shipment value)

Number of destination points served by your establishment:

Outbound Shipments during Normal Season **Duration of Normal Season:**

Shipment	Mode	Distance	Weight	Volume	Freq	Value	Transport time	Inventory storage time	Product cycle time	Transport cost	Other logistics cost
Shipment 1											
Shipment 2											
Shipment 3											
Shipment 4											
Shipment 5											

Outbound Shipments during Peak Season **Duration of Peak Season:**

Shipment	Mode	Distance	Weight	Volume	Freq	Value	Transport time	Inventory storage time	Product cycle time	Transport cost	Other logistics cost
Shipment 1											
Shipment 2											
Shipment 3											
Shipment 4											
Shipment 5											

3) Please provide the following details pertaining to five typical inbound shipments originating from your establishment during normal season and peak season.

Inbound Shipments during Normal Season

Shipment	Mode	Distance	Weight	Volume	Freq	Value	Transport time	Inventory storage time	Product cycle time	Transport cost	Other logistics cost
Shipment 1											
Shipment 2											
Shipment 3											
Shipment 4											
Shipment 5											

Inbound Shipments during Peak Season

Shipment	Mode	Distance	Weight	Volume	Freq	Value	Transport time	Inventory storage time	Product cycle time	Transport cost	Other logistics cost
Shipment 1											
Shipment 2											
Shipment 3											
Shipment 4											
Shipment 5											

Appendix B: Actual Questionnaire Used for the Survey

The 2005 Survey of Business Transportation Needs

Please answer these three questions for the primary commodities that you ship. An example survey for a hypothetical company is attached and the survey takes on average ten minutes to complete.

1. What is your choice of mode? ________________

☐ Truck ☐ Rail ☐ Both Truck and Rail

2. When choosing between truck and rail, which factors are most important?

(Please distribute 100 points among these factors based on their importance)

Factor	Points
Travel time	
On time performance	
Transportation costs	
Other logistics costs*	
Ability to track the shipment status	
Availability of special equipment to handle the shipment	
Risk of loss or damage	
Geographic coverage	
Other (specify)	
Other (specify)	
Total Points	**100**

*Other logistics costs are order processing, product handling, storage, and stock-out due to late shipments

3. ***Imagine* your shipment had to travel 200, 500, or 1,000 miles by truck and then imagine it had to travel these same distances by rail. Under each scenario, how would travel time, on time performance, transportation costs, and other logistics costs be affected? Please provide your best estimate in the table below.**

Scenario / Factor	Truck			Rail		
	200 miles	500 miles	1,000 miles	200 miles	500 miles	1,000 miles
Travel time (in days)						
Travel time reliability (Percent of shipments arriving on time)						
Transportation costs (as percent of shipment value)						
Other logistics costs (as percent of shipment value)						

Comments:

Example Response for the 2005 Survey of Business Transportation Needs

1. What is your choice of mode?

☑ *Truck* ☐ Rail ☐ Both Truck and Rail

2. When choosing between truck and rail, which factors are most important?

(Please distribute 100 points among these factors based on their importance)

Factor	Points
Travel time	*20*
On time performance	*30*
Transportation costs	*10*
Other logistics costs	*10*
Ability to track the shipment status	*5*
Availability of special equipment to handle the shipment	*0*
Risk of loss or damage	*0*
Geographic coverage	
Other (specify) Location of growth	*25*
Other (specify)	*0*
Total Points	**100**

*Other logistics costs are order processing, product handling, storage, and stock-out due to late shipments

3. ***Imagine* your shipment had to travel 200, 500, or 1,000 miles by truck and then imagine it had to travel these same distances by rail. Under each scenario, how would travel time, on time performance, transportation costs, and other logistics costs be affected? Please provide your best estimate in the table below.**

Scenario / Factor	Truck			Rail		
	200 miles	500 miles	1,000 miles	200 miles	500 miles	1,000 miles
Travel time (in days)	*¼ day*	*½ day*	*1 day*	*2 days*	*4 days*	*7 days*
Travel time reliability (Percent of shipments arriving on time)	*98%*	*98%*	*95%*	*50%*	*50%*	*50%*
Transportation costs (as percent of shipment value)	*10%*	*15%*	*20%*	*5%*	*6%*	*7%*
Other logistics costs (as percent of shipment value)	*7%*	*7%*	*7%*	*6%*	*7%*	*8%*

Comments:

Appendix C: Sample 'R' Codes Used for Modeling

R-Code for Logit Model I

```
##Working Directory###
setwd("C:/Vidya/thesis/choice-model") # set a working directory
getwd() # working directory

###Data Input###
data <- read.table("data-set-2.txt", header=TRUE) #reads data into data
#data #displays contents of data
names(data)  #displays variables in data
data1 <- data.frame(data[,-c(4)],choice = as.factor(data$choice)) #
Making Choice into a categorical variable
summary(data1)

###Input Test Set###
data2 <- read.table("test-set-2.txt", header=TRUE) #reads data into
data1
names(data2)
testdata <- data.frame(data2[,-c(7,8)],choice =
as.factor(data2$choice)) # Making Choice into a categorical variable
summary(testdata)


###Genarilized Linear Model##Logit Model-Full###
mode.glm1 <- glm(choice ~ distance + value + flow, data = data1[,],
family = binomial)
summary(mode.glm1, cor = F) #displays results of the above logit model
drop1(mode.glm1, test = "Chisq") #drops one variable at a time

####Tests####
mode.null <- glm(choice ~ 1, data = data1[1:850,], family = binomial)
#model with only the constant term
anova(mode.null, mode.glm1, test = 'Chi') #anova test on full model vs
reduced model

###Predictions on training set and accuracy of classifications
train.glm1.pred <- predict(mode.glm1, newdata = data1[], type =
'response')
train.glm1.correct <- (data1[,4]==1) == (train.glm1.pred > 0.75)
sum(train.glm1.correct)
train.glm1.railcorrect <- (data1[,4]==0) & (train.glm1.pred < 0.75)
sum(train.glm1.railcorrect)
train.glm1.truckcorrect <- (data1[,4]==1) & (train.glm1.pred > 0.75)
sum(train.glm1.truckcorrect)


###Predictions on test data set and accuracy of classifications
test.glm1.pred <- predict(mode.glm1, newdata = testdata[], type =
'response')
test.glm1.pred$class
write.table(test.glm1.pred,file="results2.txt")
test.glm1.correct <- (testdata[,7]==1) == (test.glm1.pred > 0.75)
```

```
sum(test.glm1.correct)
test.glm1.railcorrect <- (testdata[,7]==0) & (test.glm1.pred < 0.75)
sum(test.glm1.railcorrect)
test.glm1.truckcorrect <- (testdata[,7]==1) & (test.glm1.pred > 0.75)
sum(test.glm1.truckcorrect)
```

R-Code for Logit Model II

```
##Working Directory###
setwd("C:/Vidya/thesis/choice-model") # set a working directory
getwd() # working directory

###Data Input###
data <- read.table("data-set-3.txt", header=TRUE) #reads data into
data1
#data1 #displays contents of data1
names(data)  #displays variables in data1
data1 <- data.frame(data[,-c(8,9)],choice =
as.factor(data$choice),perishable = as.factor(data$perishable)) #
Making Choice into a categorical variable
summary(data1)

###Input Test Set###
data2 <- read.table("test-set-3.txt", header=TRUE) #reads data into
data1
names(data2)
testdata <- data.frame(data2[,-c(11,12)],choice =
as.factor(data2$choice),perishable = as.factor(data2$perishable)) #
Making Choice into a categorical variable
summary(testdata)

###Genarilized Linear Model##Logit Model-Full###
mode.glm1 <- glm(choice ~ value + flow+diff.tt+diff.rel+cost.diff, data
= data1[,], family = binomial)
summary(mode.glm1, cor = F) #displays results of the above logit model
drop1(mode.glm1, test = "Chisq") #drops one variable at a time

###Genarilized Linear Model##Final Model###
mode.glm3 <- glm(choice ~ value + flow+diff.tt+cost.diff, data =
data1[,], family = binomial)
summary(mode.glm3, cor = F) #displays results of the above logit model
drop1(mode.glm3, test = "Chisq") #drops one variable at a time

####Tests####
mode.null <- glm(choice ~ 1, data = data1[,], family = binomial) #model
with only the constant term
anova(mode.null, mode.glm1, test = 'Chi') #anova test on full model vs
reduced model

###Predictions on training data set and accuracy of classifications
train.glm1.pred <- predict(mode.glm1, newdata = data1[,], type =
'response')
write.table(train.glm1.pred,file="results3.txt")
train.glm1.correct <- (data1[,8]==1) == (train.glm1.pred > 0.75)
```

```
sum(train.glm1.correct)
train.glm1.railcorrect <- (data1[,8]==0) & (train.glm1.pred < 0.75)
sum(train.glm1.railcorrect)
train.glm1.truckcorrect <- (data1[,8]==1) & (train.glm1.pred > 0.75)
sum(train.glm1.truckcorrect)

train.glm3.pred <- predict(mode.glm3, newdata = data1[,], type =
'response')
write.table(train.glm3.pred,file="results3.txt")
train.glm3.correct <- (data1[,8]==1) == (train.glm3.pred > 0.75)
sum(train.glm3.correct)
train.glm3.railcorrect <- (data1[,8]==0) & (train.glm3.pred < 0.75)
sum(train.glm3.railcorrect)
train.glm3.truckcorrect <- (data1[,8]==1) & (train.glm3.pred > 0.75)
sum(train.glm3.truckcorrect)

###Predictions on test data set and accuracy of classifications
test.glm1.pred <- predict(mode.glm1, newdata = testdata[], type =
'response')
write.table(test.glm1.pred,file="results3.txt")
test.glm1.correct <- (testdata[,11]==1) == (test.glm1.pred > 0.75)
sum(test.glm1.correct)
test.glm1.railcorrect <- (testdata[,11]==0) & (test.glm1.pred < 0.75)
sum(test.glm1.railcorrect)
test.glm1.truckcorrect <- (testdata[,11]==1) & (test.glm1.pred > 0.75)
sum(test.glm1.truckcorrect)


test.glm3.pred <- predict(mode.glm3, newdata = testdata[], type =
'response')
write.table(test.glm3.pred,file="results3.txt")
test.glm3.correct <- (testdata[,11]==1) == (test.glm3.pred > 0.75)
sum(test.glm3.correct)
test.glm3.railcorrect <- (testdata[,11]==0) & (test.glm3.pred < 0.75)
sum(test.glm3.railcorrect)
test.glm3.truckcorrect <- (testdata[,11]==1) & (test.glm3.pred > 0.75)
sum(test.glm3.truckcorrect)
```

R-Code for LDA Models I and II

```
##Working Directory###
setwd("C:/Vidya/thesis/choice-model") # set a working directory
getwd() # working directory

###Data Input###
data <- read.table("data-set-3.txt", header=TRUE) #reads data into
data1
#data1 #displays contents of data1
names(data)  #displays variables in data1
data1 <- data.frame(data[,-c(8,9)],choice =
as.factor(data$choice),perishable = as.factor(data$perishable)) #
Making Choice into a categorical variable
summary(data1)
```

```
###Input Test Set###
data2 <- read.table("test-set-3.txt", header=TRUE) #reads data into
data1
names(data2)
testdata <- data.frame(data2[,-c(11,12)],choice =
as.factor(data2$choice),perishable = as.factor(data2$perishable)) #
Making Choice into a categorical variable
summary(testdata)

###Linear Discriminant Analysis-Model (LDA)####Full Model####
library(MASS)
mode.lda1 <- lda(choice ~ value + flow+diff.tt+diff.rel+cost.diff, data
= data1[,])
mode.lda1
plot(mode.lda1)

mode.lda2 <- lda(choice ~ value + flow+diff.tt+cost.diff, data =
data1[,], prior=c(0.25,0.75))
mode.lda2
plot(mode.lda2)

mode.lda3 <- lda(choice ~ value + flow+distance, data = data1[,],
prior=c(0.25,0.75))
mode.lda3
plot(mode.lda3)

###Predictions on training data set and accuracy of classifications
train.lda1.pred <- predict(mode.lda1, newdata = data1[,], type =
'response')
#train.lda1.pred$class
train.lda1.correct <- (data1[,8] == train.lda1.pred$class)
sum(train.lda1.correct)
train.lda1.railcorrect <- (data1[,8]==0) & (train.lda1.pred$class==0)
sum(train.lda1.railcorrect)
train.lda1.truckcorrect <- (data1[,8]==1) & (train.lda1.pred$class==1)
sum(train.lda1.truckcorrect)

train.lda2.pred <- predict(mode.lda2, newdata = data1[,], type =
'response')
#train.lda2.pred$class
train.lda2.correct <- (data1[,8] == train.lda2.pred$class)
sum(train.lda2.correct)
train.lda2.railcorrect <- (data1[,8]==0) & (train.lda2.pred$class==0)
sum(train.lda2.railcorrect)
train.lda2.truckcorrect <- (data1[,8]==1) & (train.lda2.pred$class==1)
sum(train.lda2.truckcorrect)

train.lda3.pred <- predict(mode.lda3, newdata = data1[,], type =
'response')
#train.lda3.pred$class
train.lda3.correct <- (data1[,8] == train.lda3.pred$class)
sum(train.lda3.correct)
train.lda3.railcorrect <- (data1[,8]==0) & (train.lda3.pred$class==0)
sum(train.lda3.railcorrect)
train.lda3.truckcorrect <- (data1[,8]==1) & (train.lda3.pred$class==1)
sum(train.lda3.truckcorrect)
```

```
###Predictions on test data set and accuracy of classifications
test.lda1.pred <- predict(mode.lda1, newdata = testdata[,], type =
'response')
#test.lda1.pred$class
test.lda1.correct <- (testdata[,11] == test.lda1.pred$class)
sum(test.lda1.correct)
test.lda1.railcorrect <- (testdata[,11]==0) & (test.lda1.pred$class==0)
sum(test.lda1.railcorrect)
test.lda1.truckcorrect <- (testdata[,11]==1) &
(test.lda1.pred$class==1)
sum(test.lda1.truckcorrect)

test.lda2.pred <- predict(mode.lda2, newdata = testdata[,], type =
'response')
#test.lda2.pred$class
test.lda2.correct <- (testdata[,11] == test.lda2.pred$class)
sum(test.lda2.correct)
test.lda2.railcorrect <- (testdata[,11]==0) & (test.lda2.pred$class==0)
sum(test.lda2.railcorrect)
test.lda2.truckcorrect <- (testdata[,11]==1) &
(test.lda2.pred$class==1)
sum(test.lda2.truckcorrect)

test.lda3.pred <- predict(mode.lda3, newdata = testdata[,], type =
'response')
#test.lda3.pred$class
test.lda3.correct <- (testdata[,11] == test.lda3.pred$class)
sum(test.lda3.correct)
test.lda3.railcorrect <- (testdata[,11]==0) & (test.lda3.pred$class==0)
sum(test.lda3.railcorrect)
test.lda3.truckcorrect <- (testdata[,11]==1) &
(test.lda3.pred$class==1)
sum(test.lda3.truckcorrect)
```

R-Code for QDA Models I and II

```
##Working Directory###
setwd("C:/Vidya/thesis/choice-model") # set a working directory
getwd() # working directory

###Data Input###
data <- read.table("data-set-3.txt", header=TRUE) #reads data into
data1
#data1 #displays contents of data1
names(data)  #displays variables in data1
data1 <- data.frame(data[,-c(8,9)],choice =
as.factor(data$choice),perishable = as.factor(data$perishable)) #
Making Choice into a categorical variable
summary(data1)

###Input Test Set###
data2 <- read.table("test-set-3.txt", header=TRUE) #reads data into
data1
```

```
names(data2)
testdata <- data.frame(data2[,-c(11,12)],choice =
as.factor(data2$choice),perishable = as.factor(data2$perishable)) #
Making Choice into a categorical variable
summary(testdata)

###Quadratic Discriminant Analysis-Model (QDA)####Full Model####
library(MASS)
mode.qda1 <- qda(choice ~ value + flow+diff.tt+diff.rel+cost.diff, data
= data1[,])
mode.qda1
plot(mode.qda1)

mode.qda2 <- qda(choice ~ value + flow+diff.tt+cost.diff, data =
data1[,])
mode.qda2
plot(mode.qda2)

mode.qda3 <- qda(choice ~ value + flow+distance, data = data1[,])
mode.qda3
plot(mode.qda3)

###Predictions on training data set and accuracy of classifications
train.qda1.pred <- predict(mode.qda1, newdata = data1[,], type =
'response')
#train.qda1.pred$class
train.qda1.correct <- (data1[,8] == train.qda1.pred$class)
sum(train.qda1.correct)
train.qda1.railcorrect <- (data1[,8]==0) & (train.qda1.pred$class==0)
sum(train.qda1.railcorrect)
train.qda1.truckcorrect <- (data1[,8]==1) & (train.qda1.pred$class==1)
sum(train.qda1.truckcorrect)

train.qda2.pred <- predict(mode.qda2, newdata = data1[,], type =
'response')
#train.qda2.pred$class
train.qda2.correct <- (data1[,8] == train.qda2.pred$class)
sum(train.qda2.correct)
train.qda2.railcorrect <- (data1[,8]==0) & (train.qda2.pred$class==0)
sum(train.qda2.railcorrect)
train.qda2.truckcorrect <- (data1[,8]==1) & (train.qda2.pred$class==1)
sum(train.qda2.truckcorrect)

train.qda3.pred <- predict(mode.qda3, newdata = data1[,], type =
'response')
#train.qda3.pred$class
train.qda3.correct <- (data1[,8] == train.qda3.pred$class)
sum(train.qda3.correct)
train.qda3.railcorrect <- (data1[,8]==0) & (train.qda3.pred$class==0)
sum(train.qda3.railcorrect)
train.qda3.truckcorrect <- (data1[,8]==1) & (train.qda3.pred$class==1)
sum(train.qda3.truckcorrect)

###Predictions on test data set and accuracy of classifications
test.qda1.pred <- predict(mode.qda1, newdata = testdata[,], type =
'response')
#test.qda1.pred$class
```

```
test.qda1.correct <- (testdata[,11] == test.qda1.pred$class)
sum(test.qda1.correct)
test.qda1.railcorrect <- (testdata[,11]==0) & (test.qda1.pred$class==0)
sum(test.qda1.railcorrect)
test.qda1.truckcorrect <- (testdata[,11]==1) &
(test.qda1.pred$class==1)
sum(test.qda1.truckcorrect)

test.qda2.pred <- predict(mode.qda2, newdata = testdata[,], type =
'response')
#test.qda2.pred$class
test.qda2.correct <- (testdata[,11] == test.qda2.pred$class)
sum(test.qda2.correct)
test.qda2.railcorrect <- (testdata[,11]==0) & (test.qda2.pred$class==0)
sum(test.qda2.railcorrect)
test.qda2.truckcorrect <- (testdata[,11]==1) &
(test.qda2.pred$class==1)
sum(test.qda2.truckcorrect)

test.qda3.pred <- predict(mode.qda3, newdata = testdata[,], type =
'response')
#test.qda3.pred$class
test.qda3.correct <- (testdata[,11] == test.qda3.pred$class)
sum(test.qda3.correct)
test.qda3.railcorrect <- (testdata[,11]==0) & (test.qda3.pred$class==0)
sum(test.qda3.railcorrect)
test.qda3.truckcorrect <- (testdata[,11]==1) &
(test.qda3.pred$class==1)
sum(test.qda3.truckcorrect)
```

R-Code for Tree Classification

```
##Working Directory###
setwd("C:/Vidya/thesis/choice-model") # set a working directory
getwd() # working directory

###Data Input###
data <- read.table("data-set-3.txt", header=TRUE) #reads data into
data1
#data1 #displays contents of data1
names(data)  #displays variables in data1
data1 <- data.frame(data[,-c(8,9)],choice =
as.factor(data$choice),perishable = as.factor(data$perishable)) #
Making Choice into a categorical variable
summary(data1)

###Input Test Set###
data2 <- read.table("test-set-3.txt", header=TRUE) #reads data into
data1
names(data2)
testdata <- data.frame(data2[,-c(11,12)],choice =
as.factor(data2$choice),perishable = as.factor(data2$perishable)) #
Making Choice into a categorical variable
summary(testdata)
```

```
#Load two packages: rpart and tree
library(rpart)
library(tree)

mode.tree <- tree(choice ~ value + flow+diff.tt+cost.diff, data =
data1[,])
summary(mode.tree)
mode.tree #the full tree
#mode.tree$frame #note branch labels change

#Plot the tree, branches proportional to decrease in impurity
plot.tree(mode.tree)
title("Mode Classification Tree", cex = 2)
text(mode.tree, label = 'yval',cex = .7) # show classes at terminal
nodes

# plot(prune.tree(mode.tree)) #Tree deviance vs. size
#plot( prune.tree(mode.tree, method = 'misclass')) #Tree
miclassification vs. size
#Pruning with Cross Validation
mode.tree.cv <- cv.tree(mode.tree,, prune.tree, method = 'misclass')
mode.tree.cv
plot(mode.tree.cv)

#AIC Tree - Penalty function approach to pruning
mode.tree.aic <- prune.tree(mode.tree, k=2) # the aic selected tree
mode.tree.aic
summary(mode.tree.aic)
plot(mode.tree.aic, type = 'u')
title("AIC Mode Classification Tree", cex = 2)
text(mode.tree.aic, cex = .7, label = "yprob")

#BIC Tree
mode.tree.bic <- prune.tree(mode.tree, k=log(nrow(data1[, ])))# the bic
selected tree
mode.tree.bic
summary(mode.tree.bic)
plot(mode.tree.bic, type = 'u')
title("BIC Mode Classification Tree", cex = 2)
text(mode.tree.bic, cex = .7, label = "yprob")

####Pruning based on C.V. results
mode.tree.5 <- prune.tree(mode.tree,,best=5) #Pruned to 5-nodes based
on Mis-class vs No of nodes in mode.tree.cv
mode.tree.5
summary(mode.tree.5)
plot(mode.tree.5, type = 'u')
title("5 Node Classification Tree", cex = 2)
#text(mode.tree.5, label = 'yval', srt = 90, cex = .7)
text(mode.tree.5, label = "yprob", cex = .7)

###Predictions on training set

train.tree.pred  <- predict(mode.tree, newdata = data1[,-8], )
#train.tree.pred
train.tree.correct <- (data1[,8]==1) == (train.tree.pred[,2]>0.50)
sum(train.tree.correct)
```

```
train.tree.truckcorrect <- (data1[,8]==1) & (train.tree.pred[,2]>0.50)
sum(train.tree.truckcorrect)
train.tree.railcorrect <- (data1[,8]==0) & (train.tree.pred[,2]<=0.50)
sum(train.tree.railcorrect)

#train.tree.pred.bic  <- predict(mode.tree.bic, newdata = data1[,], )
#train.tree.pred.bic

train.tree.pred.5  <- predict(mode.tree.5, newdata = data1[,-8], )
#train.tree.pred.5
train.tree.correct.5 <- (data1[,8]==1) == (train.tree.pred.5[,2]>0.50)
sum(train.tree.correct.5)
train.tree.truckcorrect.5 <- (data1[,8]==1) &
(train.tree.pred.5[,2]>0.50)
sum(train.tree.truckcorrect.5)
train.tree.railcorrect.5 <- (data1[,8]==0) &
(train.tree.pred.5[,2]<=0.50)
sum(train.tree.railcorrect.5)


###Predictions on test set

test.tree.pred  <- predict(mode.tree, newdata = testdata[,-11], )
#test.tree.pred
test.tree.correct <- (testdata[,11]==1) == (test.tree.pred[,2]>0.50)
sum(test.tree.correct)
test.tree.truckcorrect <- (testdata[,11]==1) &
(test.tree.pred[,2]>0.50)
sum(test.tree.truckcorrect)
test.tree.railcorrect <- (testdata[,11]==0) &
(test.tree.pred[,2]<=0.50)
sum(test.tree.railcorrect)


#test.tree.pred.bic  <- predict(mode.tree.bic, newdata = testdata[,-
11], )
#test.tree.pred.bic

test.tree.pred.5  <- predict(mode.tree.5, newdata = testdata[,-11], )
#test.tree.pred.5
test.tree.correct.5 <- (testdata[,11]==1) ==
(test.tree.pred.5[,2]>0.50)
sum(test.tree.correct.5)
test.tree.truckcorrect.5 <- (testdata[,11]==1) &
(test.tree.pred.5[,2]>0.50)
sum(test.tree.truckcorrect.5)
test.tree.railcorrect.5 <- (testdata[,11]==0) &
(test.tree.pred.5[,2]<=0.50)
sum(test.tree.railcorrect.5)
```

Appendix D: 'R' Output for Trees

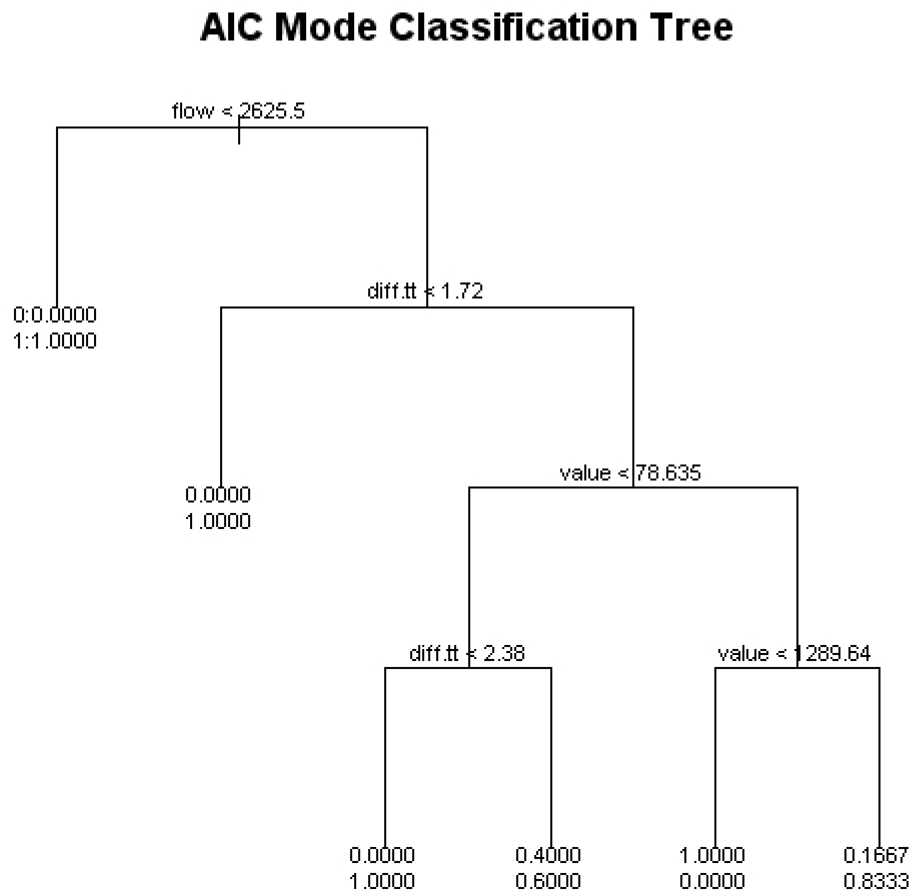
Output for Full Tree

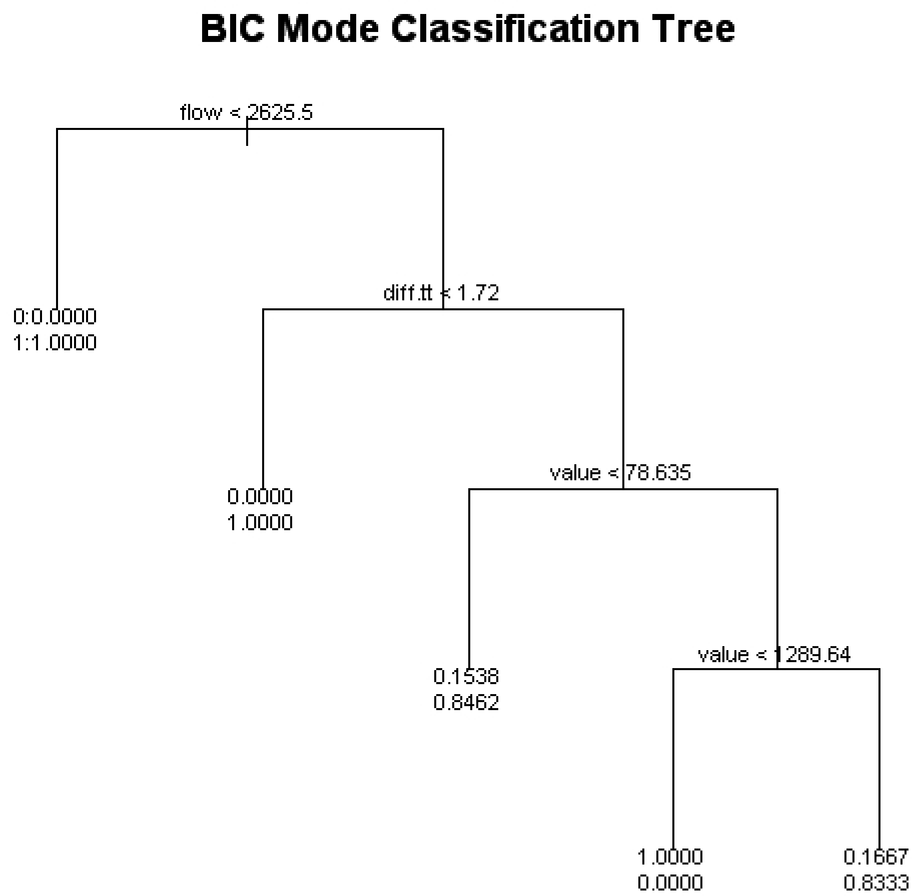
```
> mode.tree #the full tree
node), split, n, deviance, yval, (yprob)
      * denotes terminal node

 1) root 681 207.700 1 ( 0.03524 0.96476 )
   2) flow < 2625.5 626   0.000 1 ( 0.00000 1.00000 ) *
   3) flow > 2625.5 55  75.350 1 ( 0.43636 0.56364 )
     6) diff.tt < 1.72 15   0.000 1 ( 0.00000 1.00000 ) *
     7) diff.tt > 1.72 40  53.840 0 ( 0.60000 0.40000 )
      14) value < 78.635 13  11.160 1 ( 0.15385 0.84615 )
        28) diff.tt < 2.38 8   0.000 1 ( 0.00000 1.00000 ) *
        29) diff.tt > 2.38 5   6.730 1 ( 0.40000 0.60000 ) *
      15) value > 78.635 27  25.870 0 ( 0.81481 0.18519 )
        30) value < 1289.64 21   0.000 0 ( 1.00000 0.00000 ) *
        31) value > 1289.64 6   5.407 1 ( 0.16667 0.83333 ) *
```

Output for Tree Pruned using Cross Validation

```
####Pruning based on C.V. results
> mode.tree.5 <- prune.tree(mode.tree,,best=5) #Pruned to 5-nodes based on Mis-class vs
No of nodes in mode.tree.cv
> mode.tree.5
node), split, n, deviance, yval, (yprob)
      * denotes terminal node

 1) root 681 207.700 1 ( 0.03524 0.96476 )
   2) flow < 2625.5 626   0.000 1 ( 0.00000 1.00000 ) *
   3) flow > 2625.5 55  75.350 1 ( 0.43636 0.56364 )
     6) diff.tt < 1.72 15   0.000 1 ( 0.00000 1.00000 ) *
     7) diff.tt > 1.72 40  53.840 0 ( 0.60000 0.40000 )
      14) value < 78.635 13  11.160 1 ( 0.15385 0.84615 ) *
      15) value > 78.635 27  25.870 0 ( 0.81481 0.18519 )
        30) value < 1289.64 21   0.000 0 ( 1.00000 0.00000 ) *
        31) value > 1289.64 6   5.407 1 ( 0.16667 0.83333 ) *
> summary(mode.tree.5)
Classification tree:
snip.tree(tree = mode.tree, nodes = 14)
Variables actually used in tree construction:
[1] "flow"    "diff.tt" "value"
Number of terminal nodes:  5
Residual mean deviance:  0.02451 = 16.57 / 676
Misclassification error rate: 0.004405 = 3 / 681
```

Trees Pruned Using AIC and BIC Criteria




BIC Mode Classification Tree
flow < 2625.5
diff.tt < 1.72
value < 78.635
value < 1289.64
0:0.0000
1:1.0000
0.0000
1.0000
0.1538
0.8462
1.0000
0.0000
0.1667
0.8333