on Page

1. Report No.	2. Government Accession No.	3
FHWA/TX-97/1319-2		

4. Title and Subtitle	5. Report Date
FORECASTING FREIGHT TRAFFIC BETWEEN THE U.S. AND MEXICO	August 1996
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Yolanda Yulan Fang, Rob Harrison, and Hani S. Mahmassani	Research Report 1319-2
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Center for Transportation Research The University of Texas at Austin 3208 Red River, Suite 200 Austin, Texas 78705-2650	11. Contract or Grant No. Research Study 0-1319
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Transfer Office P. O. Box 5080 Austin, Texas 78763-5080	13. Type of Report and Period Covered Interim
	14. Sponsoring Agency Code

15. Supplementary Notes

Project conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.
Research project title: "Multimodal Planning and the U.S.-Mexico Free Trade Agreement"

16. Abstract

The North American Free Trade Agreement (NAFTA), a 1992 agreement negotiated by the U.S., Canada, and Mexico, has prompted new interest in freight demand forecasting. With respect to those goods moving between the U.S. and Mexico, most are transported by highway through Texas, California, New Mexico, and Arizona. Freight demand forecasting can assist transportation professionals in planning for the infrastructure maintenance required to avoid serious disruptions to trade flows across the border.

The objective of this research is to use publicly available data to develop predictive models for transport mode and Mexican destination decisions for shipments from various U.S. regions. Aggregate logit models have been calibrated for three commodities: machinery, electronics, and automobiles. A profile of Mexico and its industries is presented along with a review of past efforts in freight demand forecasting. The data set of aggregate shipments used in the model estimation is comprised of origin, destination, commodity type, mode of transport across the border, and value. Destination attributes, such as population, employment, number of firms in the industry, and number of shippers and warehouses, are also included. Based on the results of this research, origin and commodity-specific models may be used as a basis for future studies developing forecasting tools that include additional modes and commodities at a more disaggregate level.

17. Key Words	18. Distribution Statement
Multimodal operations, U.S-Mexico trade, NAFTA, freight demand forecasting, mode and destination predictions	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	109	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FORECASTING FREIGHT TRAFFIC BETWEEN THE U.S. AND MEXICO

by

Yolanda Yulan Fang
Rob Harrison
Hani S. Mahmassani

Research Report Number 1319-2

Research Project 0-1319
Multimodal Planning and the U.S.-Mexico Free Trade Agreement

conducted for the

TEXAS DEPARTMENT OF TRANSPORTATION
in cooperation with the
U.S. Department of Transportation
Federal Highway Administration

by the

CENTER FOR TRANSPORTATION RESEARCH
Bureau of Engineering Research
The University of Texas at Austin

August 1996

IMPLEMENTATION RECOMMENDATIONS

Freight demand forecasting, the subject of this report, can assist transportation professionals in planning for the infrastructure maintenance required to offset serious disruptions to NAFTA-driven U.S.-Mexico trade. Specifically, the objective of this research is to use publicly available data to develop a predictive model for transport mode and Mexican destination decisions for shipments traveling from various U.S. regions. Aggregate logit models have been calibrated for three commodities: machinery, electronics, and automobiles. A profile of Mexico and its industries is presented, along with a review of past efforts in freight demand forecasting. The data set of aggregate shipments used in the model estimation is comprised of origin, destination, commodity type, mode of transport across the border, and value. Destination attributes, such as population, employment, number of firms in the industry, and number of shippers and warehouses, are also included. Based on the results of this research, origin and commodity-specific models may be used to predict mode and destination choice. The methodology and results of this research may be applied to future studies to develop forecasting tools that include additional modes and commodities and which will be able to forecast decisions at a more disaggregate level.

This report was prepared in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

DISCLAIMERS

Rob Harrison
Research Supervisor

TABLE OF CONTENTS

SUMMARY

The North American Free Trade Agreement (NAFTA), a 1992 agreement negotiated by the U.S., Canada, and Mexico, has prompted new interest in freight demand forecasting. With respect to those goods moving between the U.S. and Mexico, most are transported by highway through Texas, California, New Mexico, and Arizona. Freight demand forecasting can assist transportation professionals in planning for the infrastructure maintenance required to avoid serious disruptions to trade flows across the border.

The objective of this research is to use publicly available data to develop a predictive model for transport mode and Mexican destination decisions for shipments from various U.S. regions. Aggregate logit models have been calibrated for three commodities: machinery, electronics, and automobiles. A profile of Mexico and its industries is presented along with a review of past efforts in freight demand forecasting. The data set of aggregate shipments used in the model estimation is comprised of origin, destination, commodity type, mode of transport across the border, and value. Destination attributes, such as population, employment, number of firms in the industry, and number of shippers and warehouses, are also included. Based on the results of this research, origin and commodity-specific models may be used to predict mode and destination choice.

CHAPTER 1. INTRODUCTION

BACKGROUND

The North American Free Trade Agreement (NAFTA), a 1992 agreement negotiated by the U.S., Canada, and Mexico, has predictably stimulated new trade among the member nations. And given the consequent demands such trade imposes on each of the countries' transportation infrastructure, NAFTA has at the same time prompted new interest in freight demand forecasting. Thus, the objective of this report is to describe how publicly available data can be used to develop a model for predicting freight transport mode; and because the relevant issue is U.S.-Mexico trade, the report also describes a method for making Mexican destination choices for shipments traveling from various U.S. regions.

Most goods moving between the U.S. and Mexico are transported by highway through Texas, California, New Mexico, and Arizona. Specifically, about 90 percent of U.S. exports to Mexico and 83 percent of U.S. imports from Mexico are transported by truck across land ports of entry (*Fiscal Notes,* 1995). Yet, as a result of Texas' and Mexico's common 1,932-km border, highways in Texas have inevitably assumed a disproportionate share of the freight traffic moving between the two countries. Thus, a major concern for transportation professionals in Texas is the need to forecast the amount of highway traffic so as to plan for the kind of infrastructure maintenance that could offset serious trade flow disruptions.

The availability of new data collected by government agencies for transportation and economic analyses has accordingly warranted an assessment of their usefulness in demand modeling. In this study, the U.S. and Mexico have been divided into regions, with aggregate logit models calibrated for three commodities: electronics, machinery, and automobiles. The methodology presented in this study seeks to facilitate future efforts in forecasting freight traffic between the U.S. and Mexico, perhaps at a more disaggregate level.

REPORT ORGANIZATION

Chapter 2 provides general background on Mexico's population, government, industries, and economic health. The information, presented for contextual purposes, is useful in understanding the forces driving the transborder movement of goods. Chapter 3 examines past efforts by researchers to develop freight demand forecasting tools. It attempts to depict some of the complexity associated with shipping decision making, focusing in particular on recent advances in technology and on the outsourcing of logistics decisions to specialized firms. It also assesses the data required for implementing these forecasting techniques.

Chapter 4 presents the structure of U.S. Bureau of Census data and outlines some additional characteristics that affect mode and destination choice. In the context of the compiled data set, Chapter 5 defines the methodological framework for devising a model useful in predicting mode and destination choice. The results of the model estimation (and their implications) are presented in Chapter 6. Chapter 7 then concludes with a summary of general findings, along with suggestions for future research in freight traffic forecasting undertaken within a binational context.

CHAPTER 2. BACKGROUND

2.1 INTRODUCTION

This chapter discusses Mexico's geography, economy, politics, infrastructure, and regional characteristics. In analyzing the patterns of trade that have evolved between the U.S. and Mexico, this chapter also discusses the nature of Mexico's maquiladora industry, as well as relevant provisions of the North American Free Trade Agreement. Because freight demand forecasting is the primary subject of this report, the following discussions are necessarily brief and are intended to provide only the contextual background for later chapters.

2.2 PROFILE OF MEXICO

2.2.1 Geography and Population

Sharing a 3,126-km border with the United States (see Figure 2.1), Mexico is topographically comprised of a central plateau lying between the Sierra Madre Oriental and Sierra Madre Occidental mountains, which meet near Mexico City and continue south to Guatemala. Being the thirteenth largest country in the world, Mexico is marked by various climatic regions, some areas being humid and tropical, others temperate or dry. As with any country having such climatic diversity, there is a mix of agriculture and industries found throughout the country.

With a total population of about 86 million, Mexico boasts a literacy rate of about 88 percent (as of 1992). Approximately 60 percent of the population is mestizo (a mix of Spanish and Mexican Indian), 30 percent is comprised of American Indian, and the remainder is Caucasian. Nearly a third of the population is employed. The most densely populated cities are the metropolitan areas of Mexico City, Guadalajara, and Monterrey.

2.2.2 Politics and the Economy

Mexico is a federal, democratic republic consisting of 31 states and the Distrito Federal. Following the 1982 peso devaluation, the Salinas administration (1988-1994) directed its efforts toward stabilizing the economy, decreasing foreign debt and inflation through price controls, and restructuring expenditures in the public sector. To encourage competition and technology transfer within industries, President Salinas introduced liberal economic policies promoting privatization of entities and deregulation of sectors. Revision of foreign investment restrictions and debt restructuring also generated economic growth and employment (Bancomext, 1993).

In December 1994, the Mexican economy was hit by another devastating peso devaluation. The current account deficit in 1994 amounted to $30 billion, resulting in a total debt of $160 billion to foreign countries (Whalen, 1995). The fall of the peso caused foreign investment in Mexico to slow and trade patterns to shift: Prior to the devaluation, northbound shipments to the U.S. were heavily outnumbered by southbound shipments to Mexico. However, as a result of the devaluation, northbound shipments diminished, leading to a more balanced bilateral trade flow.

Figure 2.1 Mexico

2.2.3 Industries

Tourism is a leading industry in Mexico. Worldwide, Mexico is the eighth most popular tourist destination, with visitors particularly attracted to its 19,159-km coastline that includes beautiful beaches and abundant aquaculture. Aside from tourism, the major industrial sectors in Mexico are agriculture, manufacturing, mining and energy, which are discussed below.

Agriculture: In 1992, this economic sector provided jobs to 30 percent of the labor force and contributed about 10 percent to the national Gross Domestic Product (GDP). Mexican agriculture produces beans, rice, wheat, barley, potatoes, maize, coffee, and fruits; the country also continues to import such U.S. bulk goods as corn, wheat, sorghum, soybeans, animal by-products, and dairy products. U.S. agricultural exports to Mexico amounted to $4 billion in 1992, with Mexico being the third largest market for U.S. beef (Winsor, 1994).

Manufacturing: Mexico's major manufactured products include automobiles, construction material, paper, chemicals, clothing and textiles, and machinery. About 19 percent of the labor force is employed in manufacturing, which comprises 33 percent of the GDP. A significant portion of the manufacturing industry is represented by maquiladora plants located along the border (and increasingly within the Mexican interior).

Mining and Energy: Mexico is the sixth-largest producer of oil in the world. In terms of oil reserves, Mexico ranks fourth, with its petroleum exports amounting to 34 percent of its foreign exchange. It is also a major importer of coal and natural gas (which relieves some of its reliance on petroleum), as well as a large producer of silver, mercury, manganese, copper, lead, and zinc (Winsor, 1994).

2.2.4 Infrastructure

The primary transport modes within Mexico include highway, rail, ship, and air, with trucks representing the predominant mode of choice. These modal choices are discussed below.

Highway: In contrast to the highly developed and intricate U.S. interstate system, Mexico's highway network is limited and poorly maintained. In 1991, only 35 percent of the approximately 240,000 km of roadway were paved (Winsor, 1994); moreover, many of the roadways that are paved are narrow and incapable of adequately carrying international truck traffic (*Texas-Mexico Multimodal Transportation,* 1993). Haulers trucking goods across the border also experience long delays as a result of complicated customs procedures (Fedorowicz et al., 1994).

Because trucks move 80 percent of all goods sent between the U.S. and Mexico, the government has focused on expanding and improving the quality of its system. For example, the Mexican Department of Communications and Transportation (Secretaria de Comunicaciones y Transportes or SCT) has allocated public funds to upgrade the highway system, and has recently instigated a program that granted concessions for the construction of over 5,000 km of highways. (The government is assuming only about 10 percent of the construction cost.) Concessions are generally transferred from the private sector back to the federal government within 20 years. Unfortunately, earlier projects called for short concession periods of about 5 years, which forced investors to charge high tolls, in turn resulting in a much lower than projected demand (*Texas-Mexico Multimodal Transportation,* 1993). The SCT has also announced a 5-year plan (1995–2000) to build 6,000 km of new highways on a concessionary basis (*U.S.-Mexico Trade and Transportation,* 1995).

Rail: The Ferrocarriles Nacionales de Mexico (FNM) is the entity charged with operating the Mexican railroad system, which was nationalized in 1937. As of 1993, railway tracks extended over 26,000 km within the country; however, performance of the system is compromised by excessive capacity and by an out-of-date operational control system (*Texas-Mexico Multimodal Transportation*, 1993).

Because of network inefficiencies relating to both freight and passenger transportation, the Salinas administration sought to modernize the rail system through various reforms. In addition to implementing computerized systems to increase rail's competitiveness with other modes, the administration also privatized rail services and equipment. In one example, Ferropuertos is providing loading and unloading services (truck and rail) for General Motors' maquiladora plants in Torreon, Coahuila. A new track joining Monterrey and Mexico, D.F., is also under construction.

Given the delays that are increasingly slowing highway operations, rail is becoming an attractive option for freight transportation. Changes in customs procedures that would allow preclearance of goods prior to arrival at the border, along with greater utilization of double-stack trains and intermodalism, have also decreased border delay (Fedorowicz, 1994). The Mexican government hopes to encourage rail modernization and, thus, attract more U.S.-Mexico freight traffic by privatizing FNM, perhaps by corridors (*U.S.-Mexico Trade and Transportation,* 1995). Partnerships between FNM and U.S. firms (with connections at ports of entry) also are enhancing modal efficiency. U.S. railway companies historically permitted access to Mexico at the border include Southern Pacific, Union Pacific, Atchison Topeka & Santa Fe, Burlington Northern, and the Texas-Mexican Railways (Winsor, 1994). The desire to capture a share of the Mexican market has led to two U.S. railroad mergers — Santa Fe with Burlington Northern (now Burlington Northern-Santa Fe) and Union Pacific with Southern Pacific (now Union Pacific).

Maritime: Maritime transit is another option being investigated by freight companies seeking to avoid border congestion. While this mode is more cost effective and energy efficient than either truck or rail, it is also the slowest mode and thus not very competitive. It is primarily used to ship such bulk, nonperishable goods as agricultural products and petroleum. Increasingly, intermodal transportation is used to transport goods more efficiently. With ongoing efforts to modernize and upgrade port facilities, containerized goods arriving at ports can now be easily transferred from ship to truck or rail.

Mexico's long coastline provides access to Long Beach, California, on the west and to the Gulf Intracoastal Waterway (which extends from Texas all the way to Florida) on the east. Mexico's major seaports (shown in Figure 2.2) include Altamira, Campeche, Ciudad del Carmen, Coatzcoalcos, Progreso, Madero, Tampico, Tuxpan, and Veracruz on the Gulf coast, and Acapulco, Ensenada, La Paz, Lazaro Cardenas, Mazatlan, Manzanillo, Puerto Vallarta, Salina Cruz, Santa Rosalia, and Topolobampo on the Pacific coast.

In the past, Mexican ports were avoided because of delays, poor cargo handling, and high probability of loss. The Mexican government privatized port operations in 1994, with the national port agency Puertos Mexicanos replaced by individual agencies known as the Integral Port Authorities. In 1995, investments of over $400 million in private sector funds were planned for infrastructure improvements and new equipment (e.g., gantry cranes) that could enhance container service and multimodal transportation.

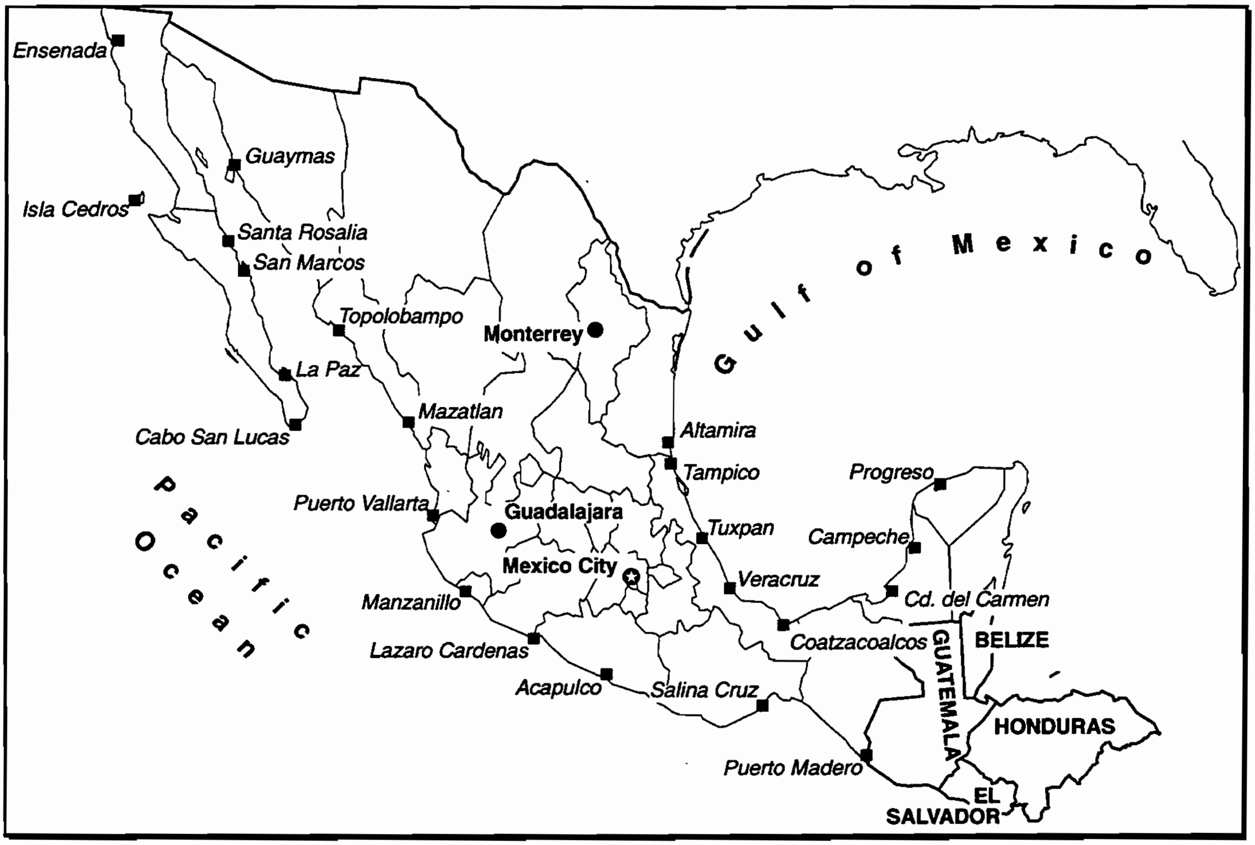


Figure 2.2 Mexican Ports

Air: Because air transport is fast but expensive, it is most often used to ship cargo that is fragile, high in value, or requires timely delivery. In 1992, cargo shipped by air totaled about 5 percent of U.S. exports to Mexico, growing at about 22 percent per year between 1987 and 1992, while U.S.-Mexico trade increased by 22 percent from 1991 to 1992 (*Texas-Mexico Multimodal Transportation*, 1993, *U.S.-Mexico Trade and Transportation*, 1995).

Mexico's air transportation network consists of 57 airports serving those Mexican cities having populations of at least 50,000, with direct flights to many U.S. cities. The system is under the national jurisdiction of the Aeropuertos y Servicios Auxiliares agency. Reforms under the previous Salinas administration (e.g., deregulation of the industry) have increased competition among carriers and have allowed financing of system improvements. The busiest airports are Mexico City, Guadalajara, Cancún, Monterrey, and Tijuana; funds have been allocated for repair and expansion of many of these airports (*Texas-Mexico Multimodal Transportation,* 1993).

2.3 MEXICAN REGIONAL PROFILES

Because of Mexico's various climates, proximity to the U.S., resources, and topography, each region possesses distinct attributes. Whereas northern Mexico is typified by dry deserts, southern Mexico is tropical (Fedorowicz et al., 1994). The three major industrial areas are Mexico City, Monterrey, and Guadalajara.

The five major regions in Mexico include the Border region, Pacific coast region, Mexico City and Vicinity, San Luis Potosi and Vicinity, and Southern Mexico, all shown in Figure 2.3. The profiles of these regions presented below (all based on the *Bancomext Trade Directory of Mexico*, 1993) will be useful in understanding the different markets and the distribution of goods entering from the U.S.

Figure 2.3 Mexican Regions

2.3.1 Border Region and Monterrey

With the exception of the Baja California peninsula, northern Mexico is dry, with mining and cattle being common industries. Given the many maquiladoras located within the northern states, this area also conducts high levels of commerce with the U.S. Almost half of the 2,000 maquiladora plants in Mexico are located in Baja California.

Forty percent of the labor force works in the maquiladora industry in Chihuahua, which is a main producer of food. Coahuila, which produces over a third of the country's steel, has recently attracted new automobile plants.

Sonora's largest sector is copper mining and its maquiladoras process food and produce electrical appliances. Just south of Sonora is Sinaloa, which exports vegetables and fish.

Monterrey, Nuevo Leon, is one of the three major industrial centers in Mexico. With its diversified, skilled labor force, it exports a wide range of goods to the U.S. Among the products produced in the many maquiladora plants are chemicals, glass, synthetic fiber, and processed foods. Tamaulipas is also characterized by a diverse economy and a skilled labor force. Along with its maquiladoras specializing in electronics and food processing, Tamaulipas has petrochemical plants and petroleum refineries. The following is a list of major companies with in-bond plants in the border region (as of 1993):

Baja California:	Fansteel, Hughes, Rockwell International, and Beckman Instruments
Sonora:	Ford
Chihuahua:	Ford, RCA, Sylvania, GE, GM, American Hospital Supply, Chrysler, Westinghouse
Coahuila:	General Motors and Chrysler
Nuevo Leon:	Alfa Group, Visa, Vitro, CYDSA

2.3.2 Pacific Coast Region and Guadalajara

Guadalajara, Jalisco, is another major technological center in Mexico. Both IBM and Hewlett Packard are located there. Its major exports include food, beverages, textiles, footwear, electronics, electrical appliances, and pharmaceuticals. The majority of maquiladoras in Guadalajara fall into the textile/apparel sector and the electronics industry (Wilson, 1992). Nayarit, Jalisco, Colima, and Guerrero all have major ports for waterborne commerce. The major industries in Guerrero and Michoacan, which have tropical climates, are agriculture and forestry. Both mainly export fruit.

2.3.3 Mexico City and Vicinity

The Federal District, the most heavily populated area in the country, serves as Mexico's political, cultural, and commercial nucleus. With extensive transportation and communications networks, Mexico City is not just a huge market for all goods, but a major location for industry as well (35,000 companies are currently located there). It is the site of about 30 percent of Mexico's manufacturing industry.

Because of their proximity to Mexico City and its extensive roadway network, the states surrounding the Federal District are also very industrialized. The state of Mexico has a large market for automobile and automotive parts industry, along with food processing, electrical and electronic equipment, textiles, and chemical engineering. Whereas Puebla manufactures automobile and automotive parts, Hidalgo specializes in producing public-transport vehicles, such as subway and railroad cars. Tlaxcala links the Port of Veracruz with Mexico City, with its major industries being chemicals and petrochemicals, textiles, and pharmaceuticals. The following lists the major companies located in each central region state:

Mexico:	Celanese Mexicana, Kimberly Clark, Industrias Resistol, Industrias Bacardi, Ford
Morelos:	BASF Vitamins, Beecham, Nissan and Pond's de Mexico
Tlaxcala:	Abbott Labs and Federal Pacific Electric of Mexico
Puebla:	Volkswagen de Mexico
Querétaro:	Kimberly Clark, General Electric, Singer, Massey Ferguson, Gerber, Kellogg's

2.3.4 San Luis Potosi and Vicinity

The economy of the central part of Mexico, just north of Mexico City, depends on mining, agriculture, and cattle raising. Durango exports marble and ranks second in Mexico in terms of quantity of gold mined. Veracruz, rich in natural resources, supports the production of coffee, rice, meat, fish, oysters, oil, and petrochemicals. Aguascalientes and Zacatecas sustain diverse industries through rail and highway connections that serve the central region. In addition to producing grapes and guava for the agricultural industry (and for export), this region boasts a burgeoning manufacturing sector. For instance, Aguascalientes produces and exports locomotives, rail parts, and electronics. The Zacatecas-Fresnillo area is encouraging production of plastics and machine parts. The following companies have established locations in these states:

Guanajuato:	Anderson Clayton, Nabisco, Campbell's, Celanese, and Motorola
Aguascalientes:	Nissan, Xerox, MORESA, Texas Instruments, General Electric

2.3.5 Southern Mexico

This region of Mexico is fairly rural and strewn with historical ruins and beaches that attract tourists. Most of the states in this area, including Oaxaca, Chiapas, Tabasco, and Campeche, have economies that depend on agriculture and livestock. Chiapas generates one-fifth of the country's electricity, while Tabasco and Campeche contain petroleum reserves that serve the manufacturing industry. Quintana Roo's economy is based largely on tourists attracted to beaches (e.g., Cancún) and to Mayan ruins; it also produces citrus fruits. Improvements at the Port of Progreso in the Yucatan have allowed international trade access to the region. Maquiladoras that produce clothing, electronics, electrical appliances, and fishing products are found in the Yucatan.

2.4 MAQUILADORA INDUSTRY

2.4.1 Background

The maquiladora industry came about as a result of Mexico's 1965 Border Industrialization Program (BIP), which sought to provide employment for Mexican workers while at the same time benefiting the economy through the use of Mexican raw materials, foreign trade, taxes, and technology transfer. The program allowed raw materials, parts, machinery, and equipment to be imported in-bond and duty free to plants located in free trade zones and within 20 km of the border

— on the condition that all goods produced would be exported. The term "in-bond" developed because the imported goods were to be eventually exported, an activity that is secured through a bond with Mexican Customs (Clement et al., 1989).

Although the assembly industry in Mexico began as a means to industrialize a particular region in Mexico (rather than as part of a national plan towards furthering its development based on exports), Mexico has since moved toward expanding the foreign trade sector (Wilson, 1992). In 1972, maquilas were permitted to locate in all interior regions of Mexico except for Mexico City, Guadalajara, and Monterrey. The Mexican government encouraged maquiladoras to locate in Mexico's interior in hopes that it would result in greater use of domestic resources in manufacturing. Restrictions on the maquiladora industry have continued to be lifted through more liberalized trade and through greater opportunities for foreign investment. For example, a portion of a maquiladora's products may be sold domestically, with a duty on the imported components sold to the Mexican market, provided these products do not compete with Mexican products (Clement et al., 1989).

2.4.2 Growth

The maquiladora industry has continued to evolve with the growth of global production sharing in which goods are assembled and produced abroad and brought back to the country of origin or exported to other markets. Given that wages are a significant component of total manufacturing cost, labor-intensive production processes are increasingly relocated to less-developed countries having lower wages. In addition, technological developments have lowered transportation and communication costs, making global production sharing an even more attractive method of production. The U.S. benefits through lower wages, transportation, and communication costs, especially considering its proximity to Mexico. U.S. facilities in border cities receive, inspect, warehouse, and export components from other areas of the country for assembly in Mexico. In 1993, there were over 2,000 maquiladora plants in Mexico, with 90 percent located in states bordering California, Arizona, New Mexico, and Texas. At that time, the industry employed nearly a half million workers in Mexico (Strong, 1996). The rapid growth of the industry has been a consequence of its ties to the U.S. economy.

But despite the benefits of transborder manufacturing processes, delays at the border — caused primarily by the need for customs documentation and inspections — are significant. Recent attempts to automate customs procedures have encouraged and facilitated commerce between the U.S. and Mexico.

2.5 NORTH AMERICAN FREE TRADE AGREEMENT (NAFTA)

2.5.1 Objectives and Provisions

The North American Free Trade Agreement (NAFTA), signed in December 1992 by the United States, Canada, and Mexico, became effective January 1, 1994. A pact to promote unrestricted movement of goods and services among these three countries, NAFTA specifies provisions that are in line with those outlined by the General Agreement on Tariffs and Trade

(GATT) for stimulating competition. To achieve a "transparent border for commerce," as described by Alan C. Courtney (1995), NAFTA was to eliminate trade barriers, promote competition with greater market access, increase investment opportunities, provide rules for resolving disputes, and protect intellectual property rights (*U.S. Custom Guide*, 1995). Some of its provisions include eliminating all tariffs and many nontariff barriers in four phases through 2008, with the majority of barriers lifted within 10 years. Nontariff barriers include quotas, import licenses, and technical standards. New investment opportunities have arisen through the agreement's less restrictive conditions for foreign ownership of businesses.

2.5.2 Implications

The elimination of tariffs and trade barriers, along with the institution of new trade incentives, has facilitated commerce within those industries relating to automobiles, machinery, and electronics. Indeed, such auto companies as Ford, Chrysler, and General Motors now have manufacturing facilities located in Mexico (Fedorowicz et al., 1994). The electronics sector has also grown as a result of the maquiladora's lower assembly costs.

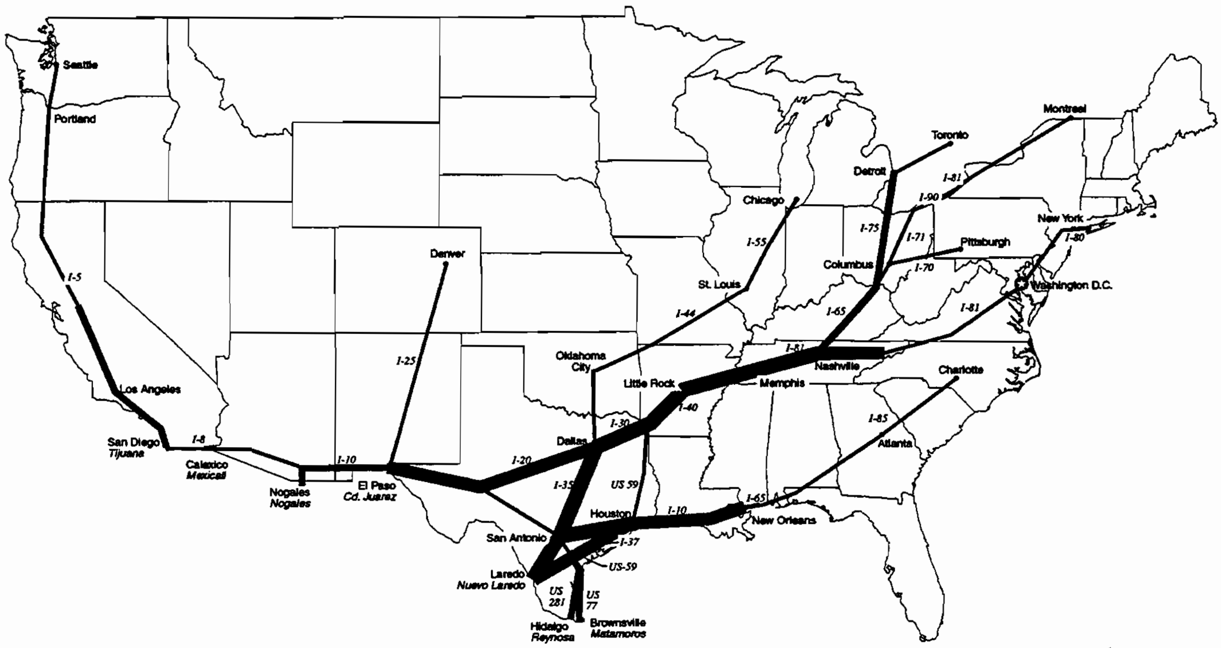
The implementation of the next phase of NAFTA, which is to permit both countries' freight companies unrestricted access to each other's highways in the contiguous border states, remains uncertain (as of June 1997). That implementation, originally scheduled to go into effect in December of 1995, has been postponed by President Clinton, while concerns of safety, illegal drugs, and infrastructure damage are fully addressed.

2.6 U.S.-MEXICO TRANSPORTATION CORRIDORS

According to the U.S. Department of Commerce's *Annual Statistics* of 1995, U.S.-Mexico trade totaled over $100 billion in imports and exports in 1994. In terms of trade value, the principal U.S. exports (for 1993) included electrical machinery, mechanical machinery, and transport equipment. Trade between the U.S. and Mexico moves between production and consumption locations by highway, rail, air, sea, and pipeline (petroleum products and natural gas). Because the scope of this project encompasses truck and rail modes of transportation, the following discussion will focus primarily on surface transportation corridors. The description of these corridors has been obtained from the Lyndon B. Johnson School of Public Affairs Policy Research Project Report No. 113, *U.S.-Mexico Trade and Transportation: Corridors, Logistics Practices and Multimodal Partnerships* (1995).

2.6.1 U.S. Surface Transportation Corridors

Highway and rail transportation corridors in the U.S. have developed in four regions of the country: west, midwest, northeast, and southeast (shown in Figure 2.4). The western corridor begins in Seattle, Washington, and runs south along I-5 through California, and continues from San Diego to Tijuana, Mexico. Continuing on I-8, the western corridor then extends east into Mexicali, Baja California, and then to Tuscon, Arizona, for access to Nogales, Sonora. Tijuana, Mexicali, and Nogales then join with Mexico's Pacific corridor. Travel along I-10 to El Paso permits access to the Mexican border city of Ciudad Juarez, which connects to the Chihuahua corridor. Rail transportation along the western corridor is provided by two major Class I railway companies: Burlington Northern-Santa Fe Railroad Company (a merger of the Atchison, Topeka, and Santa Fe Railway Company and the Burlington Northern Railroad Company), and the Union Pacific Railroad Company (a merger of the Southern Pacific Lines and Union Pacific Railroad Company).



Source: Adapted from information provided by McCray Research, San Antonio, Texas

Figure 2.4 U.S. Transportation Corridors Connecting with Mexico

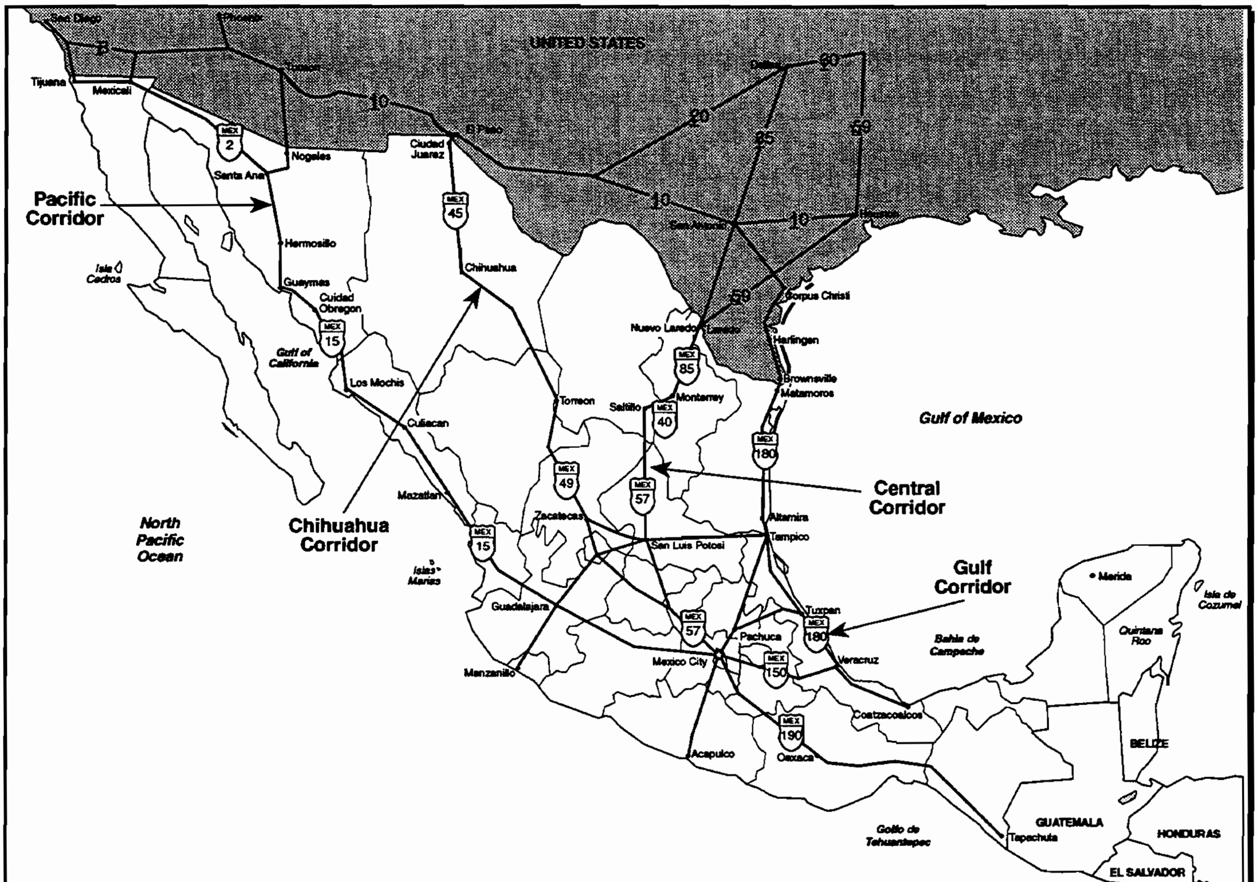
The midwestern corridor begins in Chicago, Illinois, and runs south via I-55 to St. Louis, Missouri, and then via I-44 to Oklahoma City, Oklahoma. There, I-35 links with Dallas, from which goods can travel west to El Paso, south through San Antonio to Laredo, Hidalgo, or to Brownsville. Laredo connects with Mexico's central corridor and Brownsville with Mexico's gulf corridor. The ATSF, BN, and UP have facilities along the midwestern corridor.

The northeastern corridor originates in Toronto, Montreal, and New York City, and then moves southwest to Nashville, Tennessee, and then to Texas. The Toronto leg runs south through Detroit, Michigan, to Columbus, Ohio, via I-75; there, it connects with I-65 to Nashville. The Montreal leg consists of the highway links I-81, I-90, I-71, and I-65 to Nashville. Trade from New York City travels along I-80 and connects to I-81 in Nashville. From Nashville, I-40 carries goods through Memphis to Little Rock, Arkansas, into Texas. ConRail, CSX Transportation, and Norfolk Southern Corporation are the major Class I carriers in the east.

The southeastern corridor begins in Charlotte, North Carolina, and passes through Atlanta, Georgia, and Montgomery, Alabama, via I-85, and to New Orleans via I-65. From New Orleans, it continues on I-10 to Houston, San Antonio, and El Paso, where it accesses the Texas-Mexico border. CSX Transportation and Norfolk Southern Corporation provide connections to western carriers that allow access to Mexico by rail.

2.6.2 Mexican Surface Transportation Corridors

Highway and rail corridors in Mexico, shown in Figure 2.5, have developed as a result of the country's demography and topography. The four corridors, which generally run north-south, are the Pacific, Chihuahua, Central, and Gulf corridors. The Pacific corridor begins in Tijuana at Federal Highway 2, which then joins with Federal Highway 15 from Nogales, in Santa Anna, Sonora, and continues to Mexico City. FNM operates lines that parallel both these highways to Mexico City.



Source: Adapted from information provided by McCray Research, San Antonio, Texas

Figure 2.5 Mexican Transportation Corridors with the U.S.

The Chihuahua corridor begins in Ciudad Juarez and passes through Chihuahua, Torreon, Zacatecas, and San Luis Potosi, at which point it joins with the central corridor to terminate in Mexico City. This corridor connects with the Pacific corridor at Guadalajara. The FNM rail line follows the Chihuahua corridor to Mexico City, while diverging westward of the San Luis Potosi-Mexico City segment.

The Central corridor originates in Nuevo Laredo and, via Federal Highway 85 and 40, passes through Monterrey and Saltillo. It then connects with Federal Highway 57, which carries trade to Mexico City. An FNM rail line operates along this corridor.

The Gulf corridor runs from Matamoros through Tampico and Veracruz via Federal Highway 180, then cuts westward to Mexico City via Federal Highway 150. No rail line exists between Matamoros and Veracruz; however, FNM does have a line running between Veracruz and Mexico City.

Of all the Mexican corridors, the Chihuahua corridor transports the highest value of trade relating to maquiladora products. The Central corridor, on the other hand, transports the highest value of trade unrelated to maquiladoras. The Gulf corridor does not handle a significant portion of the total tonnage or total value of commodities moving between the U.S. and Mexico (*U.S.-Mexico Trade and Transportation*, 1995).

2.7 CONCLUSIONS

This brief overview of Mexico's infrastructure, economy, climate, topography, and natural resources has been presented so as to facilitate an understanding of present U.S.-Mexico trade flows. Because the choice of freight transportation and routing is tied intricately to production and consumption markets and infrastructure, the movement of goods should thus be examined in a binational system context. The next chapter begins the process of identifying a method for forecasting freight demand and for modeling modal choice by examining previously used methods.

CHAPTER 3. FORECASTING FREIGHT DEMAND AND MODAL CHOICE MODELS

3.1 INTRODUCTION

This chapter examines previously used methods for forecasting freight demand and for modeling modal choice. As a result of both NAFTA and the December 1994 Mexican peso devaluation, the volume of traffic entering the U.S. from Mexico through the border states has increased dramatically, straining in the process existing transportation networks and facilities. Thus, predicting how traffic will be distributed among the modes can assist in maintaining the efficiency of the border transportation infrastructure.

Even before NAFTA, freight logistics practices had been evolving toward a more efficient system. New technologies and services, as well as a heightened interest in intermodalism, have allowed shippers to streamline their logistics processes and to thus reduce costs. These practices are in turn affecting how decision makers, from shippers to inventory managers, select modes (Winston, 1983). Modal choice is based on criteria such as reliability, transit time, and rates charged.

Freight forecasting techniques have been developed to determine the implications of policy and system changes on freight demand and modal choice. Early research concentrated primarily on determining which factors influenced modal decisions. These factors may be categorized into four groups (TRB, 1977):

commodity:	type, value, weight, shipment size, annual tonnage
transport system:	distance, transit time
shipper:	reliability, transport cost, frequency of service
market:	origin and destination locations, production and consumption volumes.

Research conducted during the mid-1960s and early 1970s identified factors that affected the accuracy of freight demand models, in particular level of aggregation. In addition, the lack of available disaggregated data, specifically traffic flow data which reflect modal and commodity attributes, also hampered the application and development of these models. One report — National Cooperative Highway Research Program (NCHRP) Report No. 177 (TRB, 1977) — identified the data needed for planning, and provided an extensive list of secondary sources and the data available through those sources. According to the NCHRP study, the greatest data needs include commodity flow and traffic flow data, routing data, rates/tariffs data, transport level-of-service, and unit cost data (capital and operating). Other issues that have impeded research in this area include the difficulty of aggregating commodities into homogeneous groups, and the difficulty of incorporating all attributes affecting the modal decision.

As research has progressed in the area of freight demand forecasting, some researchers have attempted to account for the total cost of a product, from its production to its distribution —

an exercise that assumes total cost ultimately affects logistics decisions (including modal choice). Nash and Whiteing (1987) asserted that modal choice is an "investment decision." This total distribution cost approach attempts to optimize the entire logistics process by minimizing the total cost of the system. Thus, the mode having the least cost is not always selected, since the cheapest mode may contribute to higher costs in other areas of the distribution system. For example, depending on the commodity, modal choice may be determined along with other logistics decisions. The decision may be affected by location of warehouses and availability of storage space (Nash and Whiteing, 1987).

The following introduces methods of forecasting modal choice and specifies the data requirements for the various models. As indicated, the type of model chosen by the analyst depends on whether the level of analysis is on a regional level, where more aggregate results are desired, or on a micro-level, where link flows by mode (i.e., more detail and less aggregation) are desired. In addition, conclusions drawn from the literature review are made regarding our effort to develop a forecasting model.

3.2 MODE SPLIT AND FREIGHT DEMAND MODELS

This section summarizes different approaches to determining freight modal choice, including some procedures that have been used to forecast freight demand. The models may be categorized as econometric or network-based models. Econometric models describe the relationship between modal choice and the factors influencing that decision, such as level of service of the mode, transit cost, and transit time. Network-based models include those that describe and simulate the transportation system.

Econometric models, which utilize cross-sectional and/or time-series data, do not require a detailed representation of the transportation system network in order to model behavior. Instead, these models attempt to determine how freight transportation service characteristics affect demand. The demand for the particular mode is dependent on its cost and on the level of service it provides. Freight demand model analyses may be further classified as either aggregate, which models average behavior of a group of decision makers, or disaggregate, which is based on the individual decision maker's behavior.

Network-based models use a representation of the transportation system to simulate and predict freight flows. From the volume of goods generated in a region, the flow of these goods may be distributed by mode throughout the network depending on the demand for the good in another region. These models, which include the spatial price equilibrium and freight network equilibrium models, require that the transportation system network be precisely detailed. Thus, the system is represented by nodes for facilities (trans-shipment nodes that do not produce or consume the commodity but rather serve as transfer points for goods), and by arcs (the infrastructure linking the nodes). Regions are then joined by a series of these links. Associated with nodes and arcs are their levels of service, which may be defined as being dependent on flows (Harker, 1987).

The spatial price equilibrium models are applied at the tactical (short-term) level, since the investment in the system is constant. Computation of an altered network is involved so the models have not been used for predictive purposes in the context of industry changes. The spatial price equilibrium model illustrates the role of transportation firms as cost functions and does not account for the decisions made by carriers (Harker, 1987).

This type of model determines network flows based on the equilibration between consuming and producing regions. The flows are assigned between regions depending on the cost of the commodity and the transport cost. The lower cost route is assigned the commodity flow. Thus, transportation demands are derived from the region's demand for the commodity (Harker, 1987).

Freight network equilibrium models describe the transportation system network and the relationship between shippers, carriers, and potential carriers. Some of the models are based on Wardrop's first and second principles, i.e., user and social equilibrium. User equilibrium is defined as the condition where all routes with flows between an origin and destination have equal, minimal costs, and those without flows have equal or higher costs. Social or system equilibrium may be defined as the condition where the total cost on the network is minimized (Ortuzar and Willumsen, 1994). These models, however, do not consider that transportation demand is derived from commodity demand (Harker, 1987).

Network equilibrium models assign flows to a network, allowing for multimodal transport of various commodities. In some previous applications, modal choice and transportation demand have been assumed to be exogenous (i.e., independent variables); however, in an attempt to predict modal split, they may be considered endogenous with the inclusion of econometric models (Guélat et al., 1990).

The remainder of this section presents mode choice and freight demand models classified by their structure and by the nature of the data used in the estimation. Aggregate models use information regarding total flows by modes and average attributes of a region, whereas disaggregate models utilize individual shipment data. Aggregate and disaggregate econometric models are presented first, followed by network-based models.

3.2.1 Econometric Models: Aggregate

Regression (Linear Probability) Models: Regression models attempt to capture any causal or correlative relationship existing between the independent variables and the dependent variable. Advantages of this model include the ability to assess the power of independent variables in explaining the dependent variable (i.e., the probability of choosing a mode). Although these models are applicable to statewide planning, there exists an inherent statistical problem of the probability possibly being outside the acceptable range of 0 to 1. Thus, the logit or probit models, which do not have this problem, are sometimes preferred. The general form of a regression model (NCHRP #177, 1977) is:

$$P_k(ij) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \mu_i$$

where

$P_k(ij)$ = estimate of the dependent variable, which is the probability of selecting mode k for the shipment between origin i and destination j,

$X_1 \ldots X_n$ = independent variables, such as transit rate or shipment size,

$\beta_0 .. \beta_n$ = parameters (i.e., coefficients), and

μ_i = error term.

A survey of relevant regression models is provided below. (NOTE: References for these models are provided at the end of each discussion.)

Perle (1964): Perle developed a regression equation to model mode split as a function of rates. He combined the data into five commodity groups and determined the volume of freight carried by truck and rail from time-series data. The independent variables included average rail and truck rates and the dummy variables: type of commodity, year, and region. The levels of service offered by each mode were not well represented. The usefulness of the results of his models were limited in that the estimated coefficients had incorrect signs. Furthermore, the dummy variables of commodity and region were more significant than the rate term in the mode split. The commodity dummy variable reflects the value of the commodity, which is correlated with rate, while the region dummy variable is a proxy for the level of service of the mode (Roberts et al., 1977).

Perle also tested a joint, aggregate demand model for truck and rail based on the same data. The decisions modeled jointly were production level and mode choice. The dependent variable was the volume of traffic carried by a mode for a commodity group in the nine regions over a year. The independent variables again were rate, region, year, and commodity. It yielded results with lower R-squared and t statistics than the previous model, primarily because this model does not include variables to reflect that transportation demand is dependent on the demand for the commodity (Roberts et al., 1977).

Mathematica (1969): This technique is an aggregate conceptual (i.e., unspecified) model system comprised of four phases and which yields the mode split of interregional commodity flows. It attempts to incorporate all shipment decisions through a sequential approach. The first stage involves predicting the total production for each of the 16 defined commodity groups using regression analysis. Then, volumes of commodities produced and received in each of the 25 regions are predicted, again using regression models, with such demographic data as income and retail sales used as explanatory variables. Next, the distribution of flows between regions are regressed against production, consumption, distance, population, and employment. Finally, the proportion of shipments by mode was determined using regressions over commodity value, average gross revenue per ton, weight, and distance. The mode choices were common carrier truck, private truck, rail, air, water, and other (Roberts et al., 1977).

Surti and Ebrahimi (1972): These researchers developed two curvilinear regression models and a linear regression model to predict modal shares for truck and rail using data obtained from the *1963 Census of Transportation*. The data published from this transportation survey were stratified into 24 commodity groups. Highway and rail were the two alternatives. The dependent variable was truck modal share in terms of total tons shipped. One of the curvilinear regression models was estimated with distance and commodity group as independent variables, whereas the other was calibrated with shipment size and commodity group. The linear regression model with shipment size and distance as explanatory variables for the 24 commodity groups resulted in higher R^2 values than the others, implying that these variables better explain the dependent variable, truck-rail traffic distribution. In addition to distance and shipment size, plant size and geographical area of the origins and destinations were also used as independent variables. However, the latter two were not very significant. Also, since commodity attributes are not included as independent variables in the regression equation, individual models must be calibrated for each commodity type (Surti and Ebrahimi, 1972).

Although Surti and Ebrahimi concluded that the linear regression model with shipment size and distance can be used to predict modal shares for carriers, destination market attributes to reflect commodity demand should be included in order to forecast freight transportation demand.

A. D. Little (1974): This model was developed to predict the proportion of freight traffic that will be assumed by barges. It incorporates both tactical and strategic variables; in other words, it considers long-term as well as short-term factors in predicting modal choice. The independent variables in this formulation include distance, circuity (water distance/rail distance), shipment size, commodity value, percentage of production and consumption facilities at the origin and destination, respectively, and dummy variables for seasonal goods and bulk goods. Circuity and distance were included as proxy variables for the level of service of barges, as compared with truck and rail (Roberts et al., 1977).

The model results revealed a strong relationship between waterborne freight movement and such strategic variables as plant location, which appeared to influence decisions to ship by water carriers more so than did tactical factors (e.g., transit time). Distance was the most important factor influencing decision making (NCHRP No. 177, 1977).

Aggregate Logit Model

Kullman (1973): Kullman estimated a binary choice logit model to predict truck and rail mode split with aggregate data. Several different combinations of variables and levels of geographic and commodity aggregations were tested. The results indicate that the commodity's value and market characteristics are important in analyzing mode choice on an aggregate level. Mode choice was determined as a function of the mode and commodity attributes. Specifically, the independent variables included highway distance, annual tonnage, commodity value, transport rates, mean transit time, and variation in transit time (Roberts et al., 1977). The explanatory variables, value per ton, tonnage, and distance, were significant at the 95 percent level and yielded the best model that still had a low R-squared (NCHRP #177, 1977).

Levin (1978): Levin (1978) estimated an aggregate logit model by ordinary least squares for market shares of truck, rail, and piggyback for 42 manufactured goods. The independent variables in his formulation were rate, transit time, and variation in transit time (a proxy for reliability). Variation in transit time had the incorrect sign; the remainder of the variables were of the correct sign and significant. He also calculated elasticities, which revealed that speed was more important than price in determining mode choice for manufactured goods. Levin found that the coefficients of shipment size, as an endogenous variable, were similar when specified as exogenous (Zlatoper and Austrian, 1989).

Murthy and Ashtakala (1987): Murthy and Ashtakala also analyzed modal choice using an aggregate logit model. Their study was conducted on a regional level (i.e., for the Canadian province, Alberta) and the data combined shipper and receiver survey responses to questions regarding commodity shipments. They organized the frequencies of data into multi-dimensional contingency tables of explanatory variables, such as average shipment size, loads (full or less than full), control (whether the respondent decided transport mode), hire (private or for hire), and type of commodity vs. modes. A log-linear model was first used to determine whether an association between the explanatory variables and the dependent variable existed. The logit model then allows the analyst to assess the combined influence of the independent variables over the dependent variable, and to determine the preference for each mode and thus modal share (Murthy and Ashtakala, 1987).

Neoclassical Economic Models

Friedlaender and Spady (1980): According to a review of freight demand models by Winston (1983), neoclassical economic models assume that the firm is a cost minimizer of production factors. The demand for transportation is derived from differentiating a cost function of shipment characteristics, factor prices excluding transportation prices, and transportation prices by mode. Friedlaender and Spady specified the short-run cost function to be a translog form. The dependent variable, expenditure share of a mode, was modeled a function of aggregate output, fixed inputs of capital and materials, and variable prices of labor, rail services, and truck services. Transportation prices included rates and inventory costs, which were affected by commodity value, density, average length of haul, and average shipment size (Zlatoper and Austrian, 1989).

The *1972 Census of Transportation* data were used to estimate the expenditure share of a mode for 5 U.S. regions and 96 manufacturing industries. Elasticities were calculated for each industry group in each region. The absolute value of the elasticities for rail were greater than 1, compared with those for truck, which were close to 1. As a result, they concluded that lower transport rates for railroads might attract greater demand for rail (Zlatoper and Austrian, 1989).

Abstract Mode Models: This approach represents the mode of transport abstractly, as a vector of values corresponding to the mode's attributes. By estimating shippers' demands for attributes (rather than the actual modes themselves), it is possible to infer what would happen if the value of the attribute changed. The model can also be applied to forecasting the demands for commodities by defining the commodities with values representing their characteristics. An advantage of this method is its capability of estimating demands for modes, which may or may not currently exist. Consequently, historical data are not needed (Terziev et al., 1975).

Although this approach was first investigated by Mathematica (1967), Herendeen also developed an abstract mode model that was used in the U.S. Department of Transportation/ National Standards of Institutes and Technology multimodal network model. His formulation is defined in NCHRP #177 (1977) as:

$$P_k(X_{ij}) = \beta_0 R_k^{\beta_1} C_k^{\beta_2} T_k^{\beta_3} F_k^{\beta_4}$$

with the constraint that $\Sigma P_k(X_{ij}) = 100$

where

$P_k(X_{ij})$ = percentage of X_{ijk} shipped by mode k,

R_k = reliability of mode k,

C_k = relative cost by mode k,

T_k = relative transit time by mode k,

F_k = relative frequency of service, and

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ = parameters determined by regression analysis.

Mathematica (1967): This method estimates freight demand for modes along each link in the transportation network. Unlike Herendeen's model, which estimates only mode choice, Mathematica developed a direct demand model that inherently utilizes the abstract mode concept. The model performs freight generation, freight distribution, and modal split in one step and yields the volume of flow between two locations by a particular mode (NCHRP #177, 1977). Because of the data requirements and the need to calibrate a model for each commodity, this approach was never tested. The model's explanatory variables are population of the origin and destination, gross regional product of the origin and destination, industrial characteristic indices, cost of each mode, transit time, and number of modes. In order to apply the model, the data must be disaggregated by mode. However, given the availability of the data, the model can predict the effect of a new mode entering the market by specifying its cost, shipping time, and other major characteristics (Terziev et al., 1975).

The direct demand model enables policy impacts such as changes in transport cost or rate to be evaluated. However, a disadvantage of abstract mode models is that the mode which is determined to be the "best" will always assume the same proportion of freight regardless of any improvements among the other modes, unless its own attributes change. Also, since it defines the modes based on their attributes, a mode may be better than another under certain circumstances but not under others. For instance, for large shipments, rail may be less expensive than truck, which may not be the case for small shipments. The same may be true for the length of haul (NCHRP #177, 1977). Furthermore, a demand model for each commodity must be estimated, since the model does not include commodity attributes (Terziev et al., 1975).

3.2.2 Econometric Models: Disaggregate

Inventory Theoretic Models: Many disaggregate econometric models, developed from the perspective of an inventory manager, allow modal decisions to occur in relation to other logistics decisions being made by the firm. Thus, these models analyze both the firm's mode and production decisions jointly. For instance, shipment size and shipment frequency are variables that can be included along with mode choice as endogenous decisions in an inventory-based model.

These decisions are made to maximize the firm's net present value of profit, or minimize total logistics costs, which is a function of the rate of production at the origin and rate of consumption at the destination.

Mathematica (1967): Mathematica developed a technique for determining mode choice based on microeconomics and inventory theory. Mode choice is determined using a shipper cost function. The explanatory variables used in the formulation are shipping costs, waiting time between shipments, transit time, and commodity type. The formulation is specified in NCHRP #177 (1977) as follows, though actual empirical calibration and testing have not been reported:

$$C_{ijk} = r_k X_{ij} + ut_k X_{ij} + a/S + WSX_{ij}/2 + h\sqrt{(S+t_k)X_{ij}}$$

where

C_{ijk} = expected total annual cost of handling shipment from origin i to destination j by mode k,

$r_k X_{ij}$ = direct shipping cost (rate * quantity shipped),

$ut_k X_{ij}$ = total in-transit carrying cost (value of time * transit time* quantity shipped),

a = cost of ordering and processing per shipment,

S = interval between shipments,

$WSX_{ij}/2$ = recipients' inventory carrying cost, and

$h\sqrt{(S+t_k)X_{ij}}$ = safety stock cost.

The selected mode minimizes the shipper's total annual cost of handling (NCHRP #177, 1977). The cost function is differentiated to obtain the marginal cost, which equals the marginal revenue at optimality. According to Terziev et al. (1975), this model is effective for evaluating a firm's operating policy or different location possibilities. Limitations to this approach include the large amount of data required to apply the model. Also, the model may not be transferable across shippers, even those having similar characteristics, since the behavior of one shipper may not explain the behavior of another shipper. Consequently, it could not be applied to regional level forecasting (NCHRP #177, 1977).

Chiang (1979): Chiang's disaggregate freight demand model considers firm logistics decisions that impact choices of mode, shipment size, and origin. He developed a short-run model based on minimizing logistics costs, such as purchase and storage costs, for a fixed demand of production inputs. The study uses substitution between factors of production and transportation as a criterion for whether the transportation problem is defined as long-run or short-run. An underlying assumption of the short-run freight demand model is that modal level of service does not affect factor substitution (Chiang, 1979).

Logistics processes were modeled in two ways: using a deterministic cost model, which accounts for all logistics costs, and using a random cost model, in which some costs are unobservable. The random cost model was estimated sequentially first for mode choice conditioned on shipment size and origin, next for shipment size conditioned on origin, and last for marginal origin choice (Chiang, 1979).

The data set used to calibrate the disaggregate, joint choice models was based on shipping documents of individual shipments from the *1972 Census of Transportation's* Shipper Survey. Annual use of the commodity, freight rate as a function of shipment characteristics, transit time, and commodity attributes (e.g., shelf life, value, and density) were also included in the data set (Chiang, 1979).

Chiang calculated demand elasticities for the joint choice of mode, shipment size, and origin, with respect to tariff charges, transit time, wait time, and percent lost or damaged. The elasticities were found to be related to commodity value and annual use of the commodity by the receiving firm. The interrelationship of mode, shipment size, and origin, captured in the joint choice model enables the shipper to respond to various policy changes by making transportation-related decisions while considering logistics costs (Chiang 1979).

McFadden, Winston, and Boersch-Supan (1985): Another study by McFadden, Winston, and Boersch-Supan (1985) focused on modeling the joint decisions of mode and shipment size for produce shipped by truck and rail. Mode choice was specified as a function of the commodity value, differences in rates and transit times, and shipment size. A dummy variable for trucks was included to capture such attributes as reliability and convenience, which are difficult to measure. Shipment size was specified as a function of transit times and the fixed and marginal rates of the modes (Zlatoper and Austrian, 1989).

The model parameters were estimated for individual choice-based data obtained from the 1977 U.S. Department of Transportation survey on produce shipments. All coefficients in the mode choice function were significant except for transit time; the coefficients for the shipment size model were not as significant. The results suggest that shipments are likely to increase and to be sent by rail with increases in truck freight rate or transit time. On the other hand, if either the fixed rate for rail or transit time increases, shipments are more likely to be transported by truck. Elasticities for mode diversion were also determined, with rail elasticities being larger than those for trucks. This result reflects the time-sensitive nature of produce shipments, which often requires them to be shipped by truck (Zlatoper and Austrian, 1989).

Intermodal Competition Model and Cross Elasticity Model: The diversion model developed for use by the American Association of Railroads (AAR) consists of the Intermodal Competition Model (ICM) and the Cross Elasticity Model (CEM). The ICM, based on a model design by Roberts (1975), assumes that shipping decisions are based on total logistics costs. It is a discrete choice model which determines rail-to-truck diversion in terms of the elasticity of the choice probability given a specified scenario. For the most part, it is used to assess the impact of truck size and weight changes on rail. The data used by the model are obtained from the Interstate Commerce Commission (ICC) Carload Waybill Sample, which provides information on a number of rail shipments between origins and destinations. The specific information from the waybill record utilized by ICM includes commodity type, origin, destination, routing, distance, equipment, railroad, and revenue. The model also requires commodity information which may be obtained from the Commodity Attribute File. This file gives such commodity characteristics as its shelf life, value, density, and special handling requirements (AAR, 1992).

The CEM, on the other hand, is a model which measures the diversion from truck to rail. It is used by Intermodal Policy Division (IPD) to determine the impact of a 10 percent reduction in railroad operating expenses on modal share. No truck traffic data similar to that used by the ICC Carload Waybill Sample are available; thus, the CEM utilizes information obtained from the National Motor Transport Data Base (NMTDB) and the ICM; the characteristics of the trucking industry are also used. These sources provide information regarding the annual mileage for all classes of trucks, truck passing counts, commodity, average rail revenue, market share, shipment size, total logistics costs, and length of haul. The methodology of the CEM is analogous to that of the ICM. That is, the total logistics costs for each mode are first calculated. This procedure is accomplished using data on market share, distance and consignment size of current truck traffic, and potential rail traffic. It then applies a logit equation to calculate modal share (Dennis, 1989).

T/RR/T Model: The Truck-Rail, Rail-Truck Diversion Model (T/RR/T) was developed by Transmode Consultants, Inc. The model uses files available on CD-ROM and requires Microsoft EXCEL 5.0 and ACCESS 1.1 for Windows. It is comprised of level-of-service models, which provide input for the mode-choice, shipment size diversion model. The diversion model is the shipper logistics cost model, developed by researchers at the Center for Transportation Studies at MIT; it is also the basis for Chiang's model and the ICM model. It is a discrete choice model used with disaggregate freight movement databases. Two versions have been formulated; one is deterministic (i.e., one mode is chosen), and the other is stochastic, which gives the probability of selecting a particular mode (Transmode Consultants, 1994).

The underlying economic theory is that each firm minimizes production costs. The lone decision maker is the shipper/receiver who chooses not only the mode of transport but also consignment size. Shippers' logistics decisions depend on annual use of commodity and minimizing logistics costs. A trade-off exists between shipping a large quantity of the good at a lower price versus having an excess of a good and having to store it (or ordering only a small shipment which costs more to transport).

The T/RR/T model utilizes three disaggregate spreadsheet databases that are accessible through a database manager: ICC Carload Waybill Sample, Rail Intermodal Sample, and Truckload Movement Sample. The model predicts modal share of rail carloads, rail intermodal trailers, rail intermodal containers, roadrailers, truckloads, longer combination vehicles, less-than-truckload (LTL) trucks, wholesalers, and private truck. It can also perform a policy impact analysis if the policy changes can be expressed in level of service and data changes (Transmode Consultants, 1994).

For each modal alternative, the total delivered cost per unit for the mode, which is comprised of total transport cost and total logistics cost, is calculated. Next, the "competitive margin" between rail and truck is assessed by comparing total delivered costs for all modes. This competitive margin comprises the quantity rail rates must be changed in order to establish competition between rail and truck. Finally, the total delivered cost per unit is recalculated to determine the best mode (Transmode Consultants, 1994).

Other Discrete Choice Models: Discrete choice models are based on random utility theory. The decision maker chooses the alternative with the highest utility compared with the other alternatives available to him/her. The utility function is an expression involving attributes of the alternatives and parameters that reflect the tastes of the decision maker. It is comprised of a systematic component, which is deterministic and captures measurable attributes, and a stochastic component, which accounts for nonquantifiable or unknown attributes. Modeling a decision maker's choice involves utilizing revealed preference data to determine these parameters, such that the choice has the highest utility (Ben-Akiva and Lerman, 1985).

Different discrete choice models have been formulated based on assumptions about the random part of the utility function, i.e., the distribution of the error terms. The logit model is based on the assumption that the error terms are logistically distributed or independently and identically Gumbel distributed (Ben-Akiva and Lerman, 1985). The multinomial logit formulation as applied to mode choice is as follows (NCHRP #177, 1977):

$$P_k(X) = \frac{e^{U(X)}}{1+e^{U(X)}}$$

where

$P_k(X)$ = probability of a shipper choosing mode k out of all the mode alternatives available to him.

$U(X)$ = $\beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n$ (utility function)

where

$X_1 \ldots X_n$ = independent variables expressed as differences, and

β_n = coefficients of X_n.

Note that several of the inventory theoretic models reviewed in the preceding subsection are also discrete choice models that have been formulated and calibrated in a random utility maximization framework.

The probit model can also be applied to analyzing freight modal split. It assumes that the disturbance terms are normally distributed and allows a more general variance-covariance matrix of the error terms. In this review, probit analysis has been applied only to modeling binary choice. However, a major shortcoming of the technique for models with more than three alternatives is that model calibration is difficult.

Discriminant analysis is another technique that may be applied to determining mode choice. Two assumptions regarding the variables are that their distributions are multivariate normal and their variance-covariance matrices are identical (NCHRP #177, 1977). If a shipper's probability distribution for each mode is known, then the value discriminant function, comprised of the linear combination of explanatory variables, is calculated to determine the mode choice. This function reflects the disutility of each mode as perceived by the shipper. Although this type of model may be applied to more than two modes, according to the study by Hartwig and Linton (1974), it is ideal for analyzing two modes.

Hartwig and Linton (1974): Hartwig and Linton tested whether the logit, probit, or discriminant model could be used to model binary freight modal choice at the disaggregate level. They assumed that the objective of the shipper is to maximize profit over all the operations of the firm. The models determine the probability of choosing the rail and truck modes as a function of commodity and modal attributes. Shipper data were obtained from waybills for rail car and truck trailers from one firm for one commodity. The independent variables of transit time, transit cost, reliability (i.e., variance of transit time), and commodity value were used in the model. For both the logit and probit models, the parameters were calculated using maximum likelihood estimation (Hartwig and Linton, 1974).

The analysis showed that all three models had statistically significant results, indicating that they are applicable to modeling freight mode choice using individual shipper data. The variables which had a significant effect on the decision were relative transit cost, reliability, and commodity value. The logit model performed best in terms of accurately predicting the shipper's mode choice. Also, elasticities and marginal rate of substitution were determined. In addition to the independent variables listed above, shipper and commodity attributes should also be incorporated into the probit model so that the model may be applied on a wider scale. The discriminant mode choice model indicated that cost and commodity value were significant in the decision (Hartwig and Linton, 1974).

Miklius (1969): Miklius developed a discriminant model to estimate freight modal split between rail and truck. The explanatory variables included in his model to determine the likelihood of a shipper choosing the rail mode were average shipment weight, distance, and plant employment at the origin, which served as a proxy variable for availability of the rail mode. Both shipment weight and transport distance were found to be significant, whereas plant employment was not. Because only one commodity was analyzed, transferability to other commodities is unlikely to be valid. The model was calibrated with aggregated data from the *1963 Census of Transportation.* This early research showed applicability of discriminant analysis to predicting mode choice (NCHRP #177, 1977).

Beuthe (1970): Beuthe developed a binary choice model that predicted the mode split between a very expensive and fast mode and a very inexpensive and slow mode. He based his discriminant model on several assumptions: one homogeneous commodity, the commodities are consumed by only one market, and all inputs to manufacturing are purchased locally. The volume of the commodity shipped by a particular mode was said to be a function of travel time, transport rate, shipment weight, and market price (Hartwig and Linton, 1974).

Antle and Haynes (1971): Antle and Haynes applied discriminant analysis to estimate freight demand for barge and rail modes. The model was calibrated with a small data set for three types of commodities obtained through shipper surveys. The data required for the model include annual tonnage between an origin and destination, distance, average travel time, shipment size, rate, difference between the rate of the chosen mode and that of the alternative mode, and handling cost. The value obtained from the function is then used to determine whether the initially chosen mode or an alternative mode will be used by the shipper. Antle and Haynes also pooled the commodities; this yielded poorer results, since commodity attributes were not represented in the model (Roberts et al., 1977). The lack of commodity attributes implies that individual models must be calibrated for each commodity type.

Winston (1981): Winston estimated probit models for mode choice using two data sets, one for rail and exempt motor carrier shipments, and the other for agricultural and industrial commodity shipments by rail, regulated motor freight, and private carriers. He assumed the decision maker was the receiver and the shipment size and firm location were fixed. The models were estimated for different commodity types. For perishable agricultural goods in the first data set, the independent variables were commodity value, shipment size, freight rate, mean and standard deviation of transit time, and reliability (i.e., coefficient of variation for transit time). In the second data set, additional variables were shipping firm location and firm sales (Zlatoper and Austrian, 1989).

The models revealed that the impact of mode level of service on mode choice was specific to the commodity. Transport mode for perishable goods, inputs for perishable goods, and goods with significant storage requirements were influenced more by level of service than by other commodities. Freight rates and firm location were found to be significant variables in determining mode choice (Zlatoper and Austrian, 1989).

From the model results, Winston calculated market demand elasticities for different mode attributes. He found that rate elasticities were greater for commodities with greater transit cost and that price was more of a factor in mode choice than service (Zlatoper and Austrian, 1989).

3.2.3 Network-Based Models

Four-Step Process for Freight (1983): This approach is analogous to the four-step process used in urban transportation planning and is detailed in NCHRP Report #260 (1983). It consists of four phases: freight generation, distribution, modal split, and traffic assignment (see NCHRP Report #260 for a full description of each step). The procedure provides different subtechniques based on the user's problem definition. Data availability may also factor into which subtechniques are chosen. The required data include base and forecast-year vehicle or commodity flows, and present and future service, cost, and rate characteristics for each mode. This procedure has been defined for rail, truck, and inland waterway transportation. Depending on the problem definition, all steps and inputs may not be required (NCHRP #260, 1983).

For the modal split step, three subtechniques have been developed. A model may be specified based on cost comparisons of marginal unit costs, rates, or physical distribution costs.

In addition, the process is flexible enough to allow other techniques to be implemented. However, since not all transportation personnel will be familiar with these other methods, they were not included as subtechniques. In each of the above subtechniques, commodity or vehicle flow data are necessary inputs. A limitation of the mode split step of this four-step methodology is that it is based on economics or logistics and does not adequately consider the service attributes of each mode (NCHRP #260, 1983).

Kresge and Roberts (1971): Kresge and Roberts developed a technique to model commodity flows on a multimodal transportation network. Roberts' transport model estimates interregional commodity flows using a gravity model and a linear program (see Kresge and Roberts, 1971, for detailed description). It consists of eight steps: commodity disaggregation, network definition, modal choice and routing, commodity distribution, commodity assignment, modal cost-performance calculation, transport price determination, and a summary of system performance measures. Commodity flows from Kresge's macroeconomic model serve as inputs into the transport model. The network consists of links representing the mode that joins supply and demand points. Transfers from one mode to another are permitted. The modal choice and routing step is based on sequential decisions made by shippers to minimize such costs as waiting time and travel time to transport a commodity. The gravity model ensures that all the demands are satisfied by producers and that there are no excess products being supplied. It provides good results for forecasting highly aggregated flows of heterogeneous commodities. The objective of the linear program is to minimize the overall cost of transporting a commodity from production points to consumption points subject to fulfilling demands. The linear program provides good results for estimating flows of homogeneous commodities such as coal and rice. After applying the linear program, the flows can be converted from value per year to tonnage. They can then be assigned to the network by minimum paths, i.e., lowest cost, and also converted to number of vehicles (Kresge and Roberts, 1971).

Guélat, Florian, Crainic (1990): This model, which predicts commodity flows on a multimodal network, is useful for strategic planning. It has been implemented in an interactive-graphic system termed the "Strategic Planning of Freight Transportation" (STAN). The network consists of links defined by the origin, destination, and mode over which commodities or passengers generate flows. The mode is specified by a cost function, type of vehicle, and capacity. Changes from one mode to another are modeled using these cost functions. The system also utilizes zonal data, such as production and consumption volumes in matrix form. The approach assigns flows to the network on the basis that commodities are shipped at a minimum total generalized cost. The behavior of shippers and carriers is also assumed to be implicit in the origin-destination matrices and modal decisions. This information, if available, may be explicitly incorporated into the matrices (Guélat et al., 1990). STAN has been used to simulate freight flows by mode for evaluating different scenarios, with modal choice being an independent variable. However, it may be possible for modal choice to be determined from econometric models, such as the ones described earlier. If modal choice is determined jointly with other decisions, the distribution of flows will have to be determined through an iterative process.

Harker (1985): Harker developed a generalized spatial price equilibrium model (GSPEM) which simultaneously determines generation, distribution, modal split, and assignment by shippers and carriers. The necessary inputs are commodity demand and price, inventory, transport costs, productivity, and carriers and their networks. The model reaches equilibrium with the balance of the purchase and sale of goods in production and consumption regions by shippers. Thus, the market demands drive transportation demand. The model represents the behavior of producers, consumers, shipper, and carriers in the equilibration process. For example, carriers are assumed to be profit-maximizers, whereas shippers determine routes based on their desire to minimize transport costs, which results in spatial price equilibrium (Harker, 1985). Tables 3.1, 3.2, and 3.3 at the end of this section summarize the variables in the econometric- and network-based models, as well as the data sources for each of these model types.

3.3 CURRENTLY AVAILABLE DATA

In order to apply these models to forecast commodity flows between the U.S. and Mexico, the data which are available need to be assessed and compared with the requirements of the model. Because these models would be used in a binational context, this section comments on data sources that might be used in model estimation. These sources include:

(1) U.S. Customs
(2) Bureau of Transportation Statistics (BTS)
(3) ICC Carload Waybill Sample
(4) NCHRP Report #178
(5) Current Lyndon B. Johnson (LBJ) School of Public Affairs Policy Research Projects
(6) Texas Department of Transportation's (TxDOT) Technology Transfer Program with Mexico, and
(7) NCHRP Project 8-30 Interim Report Draft.

The data available from U.S. Customs for the United States-Mexico Border include the following:

- mode of transport ,
- city of origin of shipment (no destination information),
- port of entry,
- ten-digit harmonized tariff code,
- value of shipment, and
- weight of gross shipment.

Additional data are also collected for water and air cargo shipments. Destination information is catalogued by a consignee number and manufacturer identification number; however, this information can only be obtained with the permission of the Internal Revenue Service (IRS). The Customs data are transmitted to the U.S. Bureau of Census for processing. No information regarding carrier or transfers to or from other modes is available.

The Bureau of Transportation Statistics (BTS) has several products available that would fulfill some of the data requirements for estimating commodity flow and mode split models. The *National Transportation Atlas Database CD-ROM* (BTS, 1996) is comprised of infrastructure and network data bases for each state and for the entire country. Although the network specifications currently include terminals, they do not show the connections between modes and terminals. The *Rail Waybill Data: 1988-1992 CD-ROM* (BTS, 1993) consists of aggregate data for shipments by rail. This source provides information on origin and destination, commodity type, tonnage, revenue, length of haul, number of cars, participating railroads, and intermodal facilities. *Waterways CD-ROM* (BTS, 1993) provides the waterways network and trade data for navigable waters. It gives 1993 domestic and foreign tonnage for major ports, as well as information on physical facilities and dredging contracts. Another source which will be available in the future is the *Commodity Flow Survey*, which gives information regarding commodity flow by mode (i.e., truck, rail, water, and air). Origin and destination, 5-digit Standard Transportation Commodity Classification (STCC) code, weight, value, and modes of transport are given for each of the shipments sampled (BTS January 1995).

The *Surface Transborder Commodity Data* represents another set of data that provides information on freight flow of U.S. imports and exports with Canada and Mexico (the information dates from April 1994). The data were gathered by U.S. Customs and provided to the Bureau of Census for processing; they were then disseminated to the public by BTS. The flows are categorized by commodity and transport mode (i.e., U.S. Postal Service — mail, rail, truck, or pipeline). The data have been sorted by both state of origin and exporter; however, caution must be exercised when using this information, since the state of origin may actually be a consolidation point for a particular commodity. Also, the exporter may not be the producer of the commodity. The files reflecting U.S.-Mexico flows contain origin and destination, commodity value, port of import or export, containerization, Harmonized Tariff Schedule/Schedule B code, and freight charges to the U.S. border for imports. Note that origins and destinations have been defined both as states and according to the 89 National Transportation Analysis Regions (NTARs), which are consolidations of the 183 Bureau of Economic Analysis Areas (BEAs) (BTS January 1995).

The ICC conducts a survey for the Association of American Railroads (AAR) of rail shipments each year for all railroad classes; this survey is dubbed the "ICC Carload Waybill Sample." The database contains commodity type by STCC code, shipment weight and number of transport units (carload, trailers, containers), origin and destination in terms of BEAs, routing (states where interchanges occur and number of interchanges), distance, equipment, railroad, and revenue. Although it is not available to the public, AAR does produce an annual Public Use Tape (NCHRP 8-30, 1993).

In addition to the data sources listed above, NCHRP Report #178, *Freight Data Requirements for Statewide Transportation Systems Planning*, provides a detailed description of various secondary sources (including their costs). This report cites sources which may be applicable to all states and contains regularly collected or recent data (at that time, 1977). Unfortunately, an obstacle to gathering data is that some carriers do not want to share data. The

NCHRP report categorizes data into five groups: traffic flow (must be collected directly, usually through shipper surveys), shipper/consignee attributes, direct/indirect impacts, carrier, and physical/operating statistics. The latter two may be obtained from secondary records, whereas information on the first two can be collected through shipper surveys (NCHRP #178, 1977).

The Policy Research Projects conducted by the LBJ School of Public Affairs may also serve as an initial inventory of information available for modeling modal choice. For instance, information exists on tonnage passing through Texas ports which may be further disaggregated into imports, exports, and domestic goods in short tons. Also, information, such as number of warehouses and access channel dimensions, as well as number of operating vessels and highway and railroad distances to cities, is available for Mexican ports. Rail service at Texas ports and railroad intermodal facilities in Texas may also be obtained. Finally, highways linking major cities, gateways, railroad, and ports is also available for Mexico and Texas (*U.S.-Mexico Trade and Transportation*, 1995). From this information, a network could be roughly depicted. However, in order to apply the network-based models described earlier, information regarding production and consumption volumes, as well as a more detailed carrier network, would be needed.

Another source of information is Mexico's transportation department, the Secretaria de Comunicaciones y Transportes (SCT). TxDOT's Technology Transfer Program with Mexico has been established to facilitate the information exchange necessary to provide an efficient system of transporting goods internationally. TxDOT has recently received diskettes from SCT which have been used in transportation planning in Mexico. SCT has implemented a geographic information system (GIS) of highways, roads, rail lines, and ports to aid in freight planning at both the national and municipal level. The GIS databases also contain socioeconomic information at both the municipal and state levels (TxDOT, 1995).

Recently, Cambridge Systematics, Inc., published an interim report of NCHRP Project 8-30, *Characteristics and Changes in Freight Transportation Demand.* The report documents the first phase of the project, which consists of survey results of various public and private groups regarding freight demand issues and ways of addressing them as well as data sources currently (in 1993) used in freight forecasts. It also identifies the key characteristics and measures of freight demand and the scope of freight databases. The following is a brief description of freight databases cited in the report which could provide the data needed for modeling:

1. TRANSEARCH: TRANSEARCH is a database developed by Reebie Associates (Greenwich, CT) that integrates truck, rail, air, and waterborne traffic from various sources. Origin and destination information is available in terms of the 183 BEAs and some Canadian provinces. Truck mode is divided into private, for-hire, and less-than-truckload (LTL). Distinction is also made between rail carload and intermodal. Commodity type by 4-digit STCC code, shipment weight, and the number of modal units are also included.

2. U.S. Imports/Exports of Merchandise on CD-ROM: This database of foreign trade is maintained by the U.S. Bureau of Census and available on CD-ROM monthly. It provides commodity type in various forms, including 10-digit Harmonized Code, Standard International Trade Classification (SITC), and Standard Industrial

Classification (SIC). Origin and destination information is limited to the country shipped to/from and the domestic district of entry or exit. Consignment value and quantity for all modes combined and value and weight for both water and air modes are given. Imports by water and air modes also have freight charges documented.

3. U.S. Exports of Domestic and Foreign Merchandise by State/Region/Port (State of Export Tapes): These magnetic tapes from the Census Bureau contain information from Shipper's Export Declaration and U.S. Customs Entry Summary and are available every 4 months. They provide commodity type (2-digit SIC and 4-digit SITC), state/region of origin, foreign country of destination, port and district of export, total value of shipment for all modes combined, and total value and weight for water and air.
4. U.S. Exports by State of Origin of Movement (MISER State of Export): Massachusetts Institute for Social and Economic Research (MISER) developed these data files from the Census Bureau's EQ912 and EA917 tapes which denote commodity by 2-digit SIC code. The data provided by these files, which may be obtained through the U.S. Department of Commerce, include state of origin and foreign country of destination, total value of shipment for all modes combined, and total weight and weight for water and air.
5. U.S. Air Freight Origin Traffic Statistics: The Colography Group (Marietta, GA) developed this data set of annual air cargo shipments (domestic and export) from surveys and trade information of the industries producing over 90% of the total air cargo shipments. Origin information in the form of state, county and Colography's designations of "market areas," which are aggregates of counties, and destination information of either domestic or foreign are noted. Also available are commodity type classified by 4-digit SIC, shipment size (express or heavy freight), annual shipment weight and value, employment and number of plants for each market area.
6. Freight Commodity Statistics: The Association of American Railroads (AAR) publishes these rail commodity statistics for all U.S. Class I railroads which are required to file a report with ICC. The data are aggregated according to Eastern and Western Districts by headquarter location and commodity type which is classified by 2- to 5-digit STCC. Total shipment weight, freight revenue and carloads by commodity are also available.
7. North American Trucking Survey (NATS): This database replaces the National Motor Transportation Database. AAR contracts Arthur D. Little, Inc., to survey drivers at various truck stops regarding the current and previous shipment. The data include commodity type (3-digit STCC), origin and destination (city, state), shipment weight, trailer type, annual vehicle-miles traveled by the driver, and carrier attributes (private, for-hire). Although this source is proprietary, it may be possible for federal and state agencies to obtain the data from AAR.
8. LTL Commodity and Market Flow Database: The American Trucking Association (ATA) contracts Martin Labbe Associates to collect data from member carriers who then are able to utilize the database. The data for LTL shipments include commodity type by service (e.g., special handling requirements), origin and destination (domestic zip codes or foreign region), distance, shipment weight, number of shipments and shipment units, and revenue.
9. Port Import/Export Reporting Service (PIERS): The Journal of Commerce maintains this database of international shipments by water from hard copy manifest reports and Customs Automated Manifest System. The data available are commodity type (6-digit Harmonized code and 7-digit PIERS code), origin and destination (shipper, city, country), U.S. and foreign port of entry or exit, shipment weight, volume and value,

and carrier and vessel names.

10. U.S. Waterborne General Imports (Exports) and Inbound (Outbound) Intransit Shipments: Records of waterborne shipments are available on tapes through the Bureau of Census. Shipments are aggregated based on commodity, ports, vessel type, and foreign country. The data are comprised of commodity type (SITC and 6-digit Harmonized code), foreign country of origin and destination, domestic and foreign ports of entry and exit, shipment weight and value, and import freight charges.
11. Waterborne Commerce and Vessel Statistics: This database, produced by the U.S. Army Corps of Engineers, relates to domestic and foreign waterborne shipments. It consists of commodity type, port tonnage summaries, domestic state of origin and destination, shipment weight, and number of vessels.
12. World Sea Trade Service: World Sea Trade Service, developed by DRI/McGraw-Hill (Lexington, MA), is comprised of both past and projected waterborne traffic for over 700 international trade routes. It identifies the commodity type by 20 SITC-based categories, origin and destination by foreign country, trade route, total shipment weight and containerloads, and number of containers.

Other freight databases are industry- or commodity-specific and provide information regarding origin and destination (production and consumption points or foreign country of import and export), shipment value, volume, and weight. Some do not distinguish between modes, while modes provided by other databases are constrained by the limitations placed on the commodity information provided for each modal route. Some of these database types are: Exports from Manufacturing Establishments, Fresh Fruit and Vegetable Shipments, Quarterly Coal Report, Natural Gas Monthly, Natural Gas Annual, Petroleum Supply Monthly, and Grain Transportation Report (NCHRP 8-30, 1993). The following tables show the data requirements for each model and the sources available for that data.

3.4 CONCLUSIONS

This chapter provided an overview of freight demand forecasting models and the findings of past researchers. It also examined current data sources that may be used to estimate models for trade flowing between the U.S. and Mexico. Some of these methods are strictly mode split models; others can forecast interregional freight flows by a particular mode with mode choice integrated within the methodology. The techniques are applicable to a range of situations that depend on the type of analysis desired.

An aggregate approach is typically chosen by the analyst for a system-level analysis of the mode share of a commodity. Firms with similar characteristics are thought to behave in a similar manner; thus, they are aggregated within a region. However, the homogeneity of a group is difficult to achieve for numerous reasons, one being that firms establish alliances with specific carriers. Another reason grouping shippers together is difficult is that their behavior is influenced by the ownership of equipment and facilities and the type of commodity being distributed. Also, behavioral differences among shippers are averaged out when aggregation is performed. Furthermore, competition among modes is not as finely depicted as on a disaggregate level (Winston, 1983).

Table 3.1. Data Requirements for Econometric Models

	DATA NEEDS											
	Commodity					Transport System		Shipper			Market	
	Type	Value	Weight	Shipment Size	Annual Tonnage	Distance	Transit Time	Cost/ Rates	Reliability	Frequency of Service	O-D (P/C)* Volumes	O-D (P/C)* locations
AGGREGATE MODELS												
Regression												
Perle (1964)	X							X				
Mathematica (1969)	X	X				X					X	
Surti and Ebrahimi (1972)[a]	X			X		X						X
A. D. Little (1974)[b]		X		X		X						X
Aggregate Logit												
Kullman (1973)		X			X	X	X	X[c]	X			
Murthy and Ashtakala (1987)	X			X								

Notes:
* O-D: origin and destination; P/C: production and consumption
a. Model also requires plant size.
b. Model also requires circuity, bulk commodity, and seasonality.
c. These variables are represented relative to competing mode.

	DATA NEEDS											
	Commodity					Transport System		Shipper			Market	
	Type	Value	Weight	Shipment Size	Annual Tonnage	Distance	Transit Time	Cost/ Rates	Reliability	Frequency of Service	O-D (P/C) Volumes	O-D (P/C) locations
DISAGGREGATE MODELS												
<u>Abstract Mode</u>												
Herendeen (1969)							X[c]	X[c]	X	X[c]		
Mathematica (1967)[d]							X	X				
<u>Linear Programming</u> Tripp (1972)[e]						X						
<u>Microeconomics and Inventory Theory</u> Mathematica (1967)[f]	X						X	X				

d. Model also requires number modes and industrial character index.
e. Model also requires: commmodity attributes (loading characteristics, susceptibility to loss/damage, traffic volume and regularity, equipment required, route characteristics), weather, and traffic density. Constraints: demand satisfaciton, capacity, logistic system.
f. Model also requires inventory costs and value of time.

	DATA NEEDS											
	Commodity						Transport System		Shipper			Market
	Type	Value	Weight	Shipment Size	Annual Tonnage	Distance	Transit Time	Cost/ Rates	Reliability	Frequency of Service	O-D (P/C) Volumes	O-D (P/C) locations
<u>Discrete Choice</u>												
Hartwig and Linton (1974)		X					X	X	X			
Miklius (1969)[g]			X			X						
Beuthe (1970)[h]			X				X	X				
Antle and Haynes (1971)[i]				X	X	X	X	X				

g. Model also requires plant employment.
h. Model also requires market price.
I. Model also requires handling cost.

Table 3.2. Econometric Models: Data Sources for Frequently Used Variables

Data Need	United States Sources*				Mexican Sources*			
	Truck	Rail	Water	Air	Truck	Rail	Water	Air
Commodity								
• Type	1, 3, 5, 6, 13, 14	1, 2, 3, 5, 6, 11, 12	1, 4, 5-9, 15-18	5-10	1, 4	1, 2	1, 6	1
• Value	1, 3, 5,	1, 2, 3, 5,	1, 5, 7, 8, 9, 15, 16	5, 7, 8, 9, 10	1	1	1	1
• Weight	1, 5, 6, 13, 14	1, 2, 5, 6, 11, 12	1, 5, 6, 7, 8, 9, 15, 16, 17	5, 6, 7, 8, 9, 10	1, 4	1, 2	1, 6	1
• Shipment Size	6	6, 12	6, 15, 18	6, 10				
• Annual Tonnage		2		10	1	1, 2	1, 6	1
Transport System								
• Distance	14	2, 11				2		
• Transit Time								
Shipper*								
• Cost/Rate								
• Reliability								
• Frequency of Service								
Market								
• Origin-Destination Volumes								
• Origin-Destination Locations*	1 (origin only), 3, 5, 13, 14	1 (origin only), 2, 3, 5	1 (origin only), 4, 5, 8, 9, 15, 16, 17, 18	5, 8, 9, 10	1, 4	1, 2	1, 6	1

* See notes on next page.

Notes: The numbers on the previous page designate the following sources from which the data may be obtained.

United States Sources

1. U.S. Customs
2. Rail Waybill Data: 1988-1992 CD-ROM
3. Surface Transborder Commodity Data Diskettes
4. Waterways CD-ROM
5. Commodity Flow Survey
6. TRANSEARCH
7. U.S. Imports/Exports of Merchandise
8. U.S. Exports of Domestic and Foreign Merchandise (State of Export Tapes)
9. U.S. Exports by State of Origin of Movement (MISER)
10. U.S. Air Freight Origin Traffic Statistics
11. ICC Carload Waybill Sample
12. Freight Commodity Statistics
13. North American Trucking Survey
14. LTL Commodity and Market Flow Database
15. Port Import/Export Reporting SErvice
16. U.S. Waterborne General and Intransit Service
17. Waterborne Commerce and Vessel Statistics
18. World Sea Trade Service

Mexican Sources

1. Secretaría de Comercio y Fomento Industrial (SECOFI)
2. Ferrocarriles Nacionales de México (FNM)
3. Instituto Nacional de Estadística, Geografía e Informática (INEGI)
4. SCT- Dirección General de Servicios Técnicos y Concesiones
5. Caminos y Puentes Federales de Ingresos y Servicios Conexos (CAPUFE)
6. SCT- Dirección General de Puertos y Marina Mercante
7. Aeropuertos y Servicios Auxiliares (ASA)

Note that the Mexican sources listed above are organizations which maintain the data rather than the name of the database as with the U.S. sources. At the time of this report, no description of the specific databases had been obtained.

Shipper information is difficult to obtain for each shipment because they do not want their identification revealed. Public use tapes have been modified so that any means of identifying the shipper are removed.

Origin and destination locations are documented differently by each source. For instance, the Surface Transborder Commodity Data has origin and destination as state or exporter location which may not be the same location as the producer of the good. Also, the locations could be transfer or storage points for a type of commodity and not the original production point. Also, for U.S. exports, the ultimate Mexican state of destination is provided; however, for U.S. imports, only the country of origin is listed.

Table 3.3. Network-Based Models: Required Inputs and Data Sources

Model	Inputs	United States Sources	Mexican Sources
Four Step Process For Freight (1983)	Base and forecast year vehicle or commodity flows Present and future mode service Rate characteristics for each mode	Commodity Flow Survey N/A N/A	
Kresge and Roberts (1971)	Commodity Flows Network Representation by origin, destination, and mode Costs such as waiting time, transit time, direct shipping cost	From Macroeconomic Model (Kresge) National Transportation Atlas CD-ROM N/A	SCT diskettes (Technology Transfer Program)
Guélat, Florian, Crainic (1990)	Network Representation by origin, destination, and mode Origin-Destination Matrices	National Transportation Atlas CD-ROM Commodity Flow Survey	SCT diskettes (Technology Transfer Program)

A disaggregate modeling approach, which analyzes individual movements, may model demand better because it attempts to reflect the rationale behind freight transportation decisions. The characteristics of the individual shipper may be captured, since the variation of the decision making group is retained, as opposed to aggregation in which variation is lost. According to Winston (1983), disaggregate analysis is more applicable to optimizing logistic processes and more accurate in estimating market elasticities. Disaggregate analysis is generally based on behavioral or inventory theory. Behavioral models, based on utility maximization, do not consider annual commodity production, seasonality, or logistics decisions such as shipment size and frequency as inventory theory models do. Shipment size and frequency decisions are based on inventory at production and consumption points. The firm maximizes its profit by optimizing mode choice, shipment size, and shipment frequency (Winston, 1983).

Disaggregate data are costly and difficult to obtain since characteristics of chosen and unchosen alternatives must be obtained. Thus, all the modes available to the firm for shipping, along with modal characteristics, need to be determined for modal split. Estimation may also be difficult if the number of modes available to the shipper is high or if many factors are considered endogenous in the model specification. Therefore, this type of analysis may not be practical for studying behavior at the regional level (Winston, 1983).

An important consideration in estimating a freight demand model for U.S.-Mexico trade is the binational platform for ever-changing policies as trade barriers are eliminated. The dynamic nature of the NAFTA environment calls for frequent analyses using the latest data. Other issues which complicate freight demand modeling include mergers among carriers, shipper alliances, changes in logistic practices, and the economic instability of Mexico. The move toward multimodal transportation suggests that shippers utilize combinations of modes to ensure cost-effective and efficient transport of goods. Containerization is also facilitating the use of multiple modes.

The predictive power of models depends largely on the quality of available data and the accurate representation of attributes influencing demand and mode choice. Various suggestions for improving the predictive ability of modal choice models and demand models can be found in literature. For instance, Winston (1983) recommends that a joint choice model of mode choice, shipment size, and frequency of shipments be developed to better reflect the firm's decision process. He also states that mode choice is related to the location of the firm and the market area which should be represented in the models. Harker (1987) believes that models may be improved with the integration of econometric and network-based models.

Using the findings of the literature review and considering the data, the present report attempts to model mode and destination choice jointly as a function of market, modal, and shipment attributes. Since the scope of the research includes two nations and the existing data to be used in model estimation are aggregated, a regional level analysis will be performed. Although the specific behavior of a firm cannot be modeled with this aggregate approach, the considerations in firm decision making will be modeled. Kullman's estimation of a binary mode choice model with aggregate data suggests that both mode and commodity attributes influence choice. Consequently, the aggregate logit model for mode and destination in this study should include these attributes such as distance and value.

The selected data are publicly accessible and periodically collected to facilitate implementation and model updating. The data set development is described in Chapter 4. Chapter 5 identifies the methodology and model structure selected for analyzing destination and mode choice of freight transportation between the U.S. and Mexico.

CHAPTER 4. DATA

4.1 INTRODUCTION

Federal efforts to assist transportation planning activities in anticipation of (and in some instances in response to) traffic changes resulting from NAFTA have prompted some reexamination of methods for forecasting freight transportation demand. Chapter 3 identified a variety of data sources available for freight demand analysis and modal choice. One such source is the data distributed by the U.S. Bureau of the Census, which are collected through shipper export declarations, the Automatic Broker Interface, and Customs entry documents for trade between the U.S., Canada, and Mexico. Census Bureau data served as the primary data source in this report. Other information contributing to the data set, such as population and employment, will also be described. Additional data, although not incorporated into this analysis, are also summarized insofar as they may be useful in future work.

4.2 SURFACE TRANSBORDER COMMODITY DATA

The Surface Transborder Commodity Data file of U.S. exports from August 1994 through July 1995 was used as the basis for the data set in this analysis of freight flows. It is comprised of the following variables:

- disaggregated method of transportation,
- Schedule B code,
- origin state,
- Mexican state of destination,
- foreign country of destination which is always Mexico,
- value,
- month and year of shipment, and
- count.

Mode of transport across the border is categorized as U.S. Postal Service (mail), truck, rail, pipeline, and "other," which includes unknown modes. However, observations of only the surface modes of truck and rail were used because they comprise the majority of all export shipments and contribute the most to the commercial traffic crossing the border.

Schedule B code is the two-digit Harmonized Schedule (HS) commodity code, a classification system for U.S. exports. The commodity groups of interest in this analysis are machinery and mechanical equipment (HS 84), electronics and electrical equipment (HS 85), and automobiles and automotive parts (HS 87).

U.S. state of origin is denoted by the U.S. mail abbreviations for the 50 states, with "DU" indicating an unknown origin state. Mexican destination states are also two-character abbreviations for the 31 states and Distrito Federal, with "OT" indicating an unknown destination state. Data are also collected with exporter state as the origin. However, neither origin nor exporter state actually represents the true production point for exports because origin state could be a consolidation point and exporter state could be the corporate headquarters and not the actual producer of the good. Because the exporter and origin states differ at times, the aggregation of shipments is different. U.S. state of origin, rather than exporter state, is used as the origin of the shipment because it almost always reflects the beginning point of some shipment, whether it be a trans-shipment location or true origin, whereas the location of the exporter may have no relation to the physical movement of the shipment.

The foreign country of shipment destination is always Mexico, which is symbolized by the code "2010." Value of the commodity is expressed in U.S. dollars. The statistical month and year indicate the time period in which the shipment crossed the border.

Count is the number of individual shipments that have been summarized into one observation. Aggregation of shipments is performed by the Census Bureau according to whether multiple occurrences of commodity shipments by the same mode, from the same origin to the same destination during the same time period exist. Thus, the individual records in the data file may not represent a single shipment and also do not reflect the number of trucks or railcars transporting the good.

4.3 SUPPLEMENTARY DATA

Other variables believed to influence the demand for the goods in the destination market have been added to the data set. In Chapter 3, distance, population, employment, and industry concentration were identified as attributes possibly affecting destination and mode choice.

4.3.1 Distance

Distances between major cities were determined for the U.S. and Mexico, since the distance reflects the cost of shipping and affects the transport mode chosen. Highway distances have been obtained from *Tripmaker* software package developed by Rand McNally. The distances were determined for the quickest (least time) rather than shortest (least distance) route between the origin and destination. The decision maker, possibly a shipper or a firm, would like to minimize his/her transport cost, yet still be provided quality service by the trucking firm. Time is another variable that affects the mode and destination choices; unfortunately, no actual time information associated with the shipments was collected. A proxy for time could be length of haul, since the shorter the distance, the quicker the trip in most cases. Decision makers must meet market demands that at times require the goods to be delivered as promptly as possible. The quickest route, which more likely implies highways capable of handling high-speed traffic, is probably safer and faster than the shortest route. Drivers may prefer these high-speed facilities. Note that the quickest route is based on the speed of the facilities and does not account for any delays that might be experienced at the border.

No commercial software for railway network analysis is known to exist for transport between both countries. The particular port of entry is unknown; however, it may be hypothesized based on the location of the destination and the commodity being shipped. For this project, however, the rail distances between U.S. origins and Mexican destinations were estimated based on the distances in the *Rand McNally Railroad Atlas*. Given the limited extent of the rail network, this approach may be adequate in determining rail distances.

Another approach to determining U.S.-Mexico distances could be implemented in future applications, although time and resource restrictions make it unfeasible at this time. *The National Transportation Atlas Data Bases* consist of both U.S. and Mexican highway and rail networks. These files are in ASCII comma delimited format and contain information on links and nodes. The U.S. railroad network is categorized by classes of railroads and by the type of service the companies provide. A program to extract the networks of the major railroads in freight transportation, such as Class I railroads (which earn the greatest revenue), must be created to make information, such as latitude and longitudinal coordinates for distance calculations, accessible by a geographical information system (GIS). The National Mexican Railroad network could be merged into the same layer with the U.S. railroad system, with distances then estimated. Currently, the Railroad Information System (a company) of Georgetown, Texas, is working on developing a computerized rail network system for the U.S. that could provide distance information. Perhaps in the future the Mexican rail system might be integrated, so that international distances could be readily determined.

4.3.2 Demographic Information

Population and employment percentage data for each U.S. and Mexican state were included in the data set because they are believed to influence the annual use of commodities. The greater the population, the greater the demand for goods. Likewise, the higher the employment rate, the higher the expenditure on goods. The latest U.S. population and percent unemployed statistics estimated for 1994 were obtained from *The World Almanac* and *Book of Facts 1996,* while Mexican population and employment by state were obtained from the 1990 Mexican Census. The number of U.S. firms by industry has been obtained for each state from County Business Patterns of 1992 according to SIC code. Since SIC and HS codes do not correspond one-to-one, the actual number of firms are approximate. This databank also provides the number of truck and warehousing establishments in each state. Unfortunately, the number of railway companies in each U.S. state is not provided by this source. Both the number of Mexican firms in the machinery, electronics, and automobile industries and the number of rail and truck shipping companies and warehouses were determined for each Mexican state using the 1993 Bancomext Trade Directory for Mexico. This directory publishes information on participating firms and thus may not include all firms in the industry in each state. Nevertheless, this attribute could serve as a proxy for industry concentration by sector in the state, which would help explain the number of related shipment types.

4.4 AGGREGATION

Because this study encompasses a large geographical region, namely, the U.S. and Mexico, the data have been aggregated from state-level shipment data into regions for destination choice model estimation purposes. With state level data, the number of choice alternatives is extremely high. It depends on the number of modes and destinations, which in this case results in 64 alternatives (two modes multiplied by 32 destinations). Thus, an aggregation of states into regions is necessary to reduce this choice set to a more manageable size. The states in both countries have been aggregated based on the major trade corridors that have developed, principal industries, manufactured goods, and crops within the state as well as geographic location. Neighboring states having low levels of international trade and similar location characteristics (e.g., coastal states with reliance on similar industries) are aggregated together.

The distances between regions have been determined based on a weighted average of distances between states within each region. The distances between states are weighted by the number of shipments between the states. Within each region, a centroid, typically a major city which does not necessarily correspond to the regional distances calculated, has been defined as an indicator of the region location. The centroid was selected based on industrialization of the city and whether major highways and/or rail terminals or intermodal locations might exist. Tables 4.1 and 4.2 show the regions, the centroids of those regions, and the states that comprise them for both the U.S. and Mexico.

Table 4.1. U.S. Origin Regions

	REGION	CENTROID	STATES
1	West	San Francisco, CA	Washington, Oregon, California
2	Mountain	Salt Lake City, UT	Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico
3	West North Central	Sioux City, IA	North Dakota, South Dakota, Kansas, Nebraska, Minnesota, Iowa, Missouri
4	West South Central	Dallas, TX	Texas, Oklahoma, Arkansas, Louisiana
5	East North Central	Chicago, IL	Wisconsin, Indiana, Illinois, Michigan, Ohio
6	East South Central	Birmingham, AL	Kentucky, Tennessee, Alabama, Mississippi
7	New England	Boston, MA	Maine, Vermont, New Hampshire, Rhode Island, Connecticut, Massachusetts
8	Mid-Atlantic	Philadelphia, PA	New York, New Jersey, Pennsylvania
9	South Atlantic	Atlanta, GA	Florida, Georgia, North Carolina, South Carolina, West Virginia, Maryland, Washington, D.C., Delaware

Table 4.2. Mexican Destination Regions

	REGION	CENTROID	STATES
1	Baja California	Tijuana	Baja California and Baja California Sur
2	Sonora/Sinaloa	Hermosillo	Sonora and Sinaloa
3	Chihuahua/Durango	Chihuahua	Chihuahua and Durango
4	Coahuila	Torreon	Coahuila
5	Nuevo Leon	Monterrey	Nuevo Leon
6	Tamaulipas	Tampico	Tamaulipas
7	North Central	San Luis Potosi	Zacatecas, Aguascalientes, San Luis Potosi, and Guanajuato
8	West		Michoacan, Colima, Jalisco, and Nayarit
9	Central	Mexico City	Mexico, Distrito Federal, Morelos, Querétaro, Puebla, Hidalgo, Tlaxcala
10	South Gulf	Merida	Quintana Roo, Yucatan, Campeche, and Tabasco
11	Veracruz	Veracruz	Veracruz
12	South Pacific	Oaxaca	Oaxaca, Chiapas, Guerrero

4.5 SUMMARY OF EXPLANATORY DATA ANALYSES

An initial analysis of the data showed that the origins exporting the highest total value of shipments include Texas, California, Michigan, Illinois, Arizona, and Ohio (Table 4.3). The high number and value of shipments attributed to the border origin states may be a result of trans-shipment or warehouse locations in these states. Electronics shipped by truck comprise the highest value of shipments from Texas, California, and Arizona, presumably because of their proximity to the Mexican border states (which house many maquiladoras). Michigan, on the other hand, ships the greatest value of vehicles by truck. The shipments of highest total value for Illinois and Ohio are machinery by truck.

The most frequent destinations in number of shipments include Chihuahua (130,270), Mexico (80,886), Baja California (73,338), Tamaulipas (63,700), Distrito Federal (59,461), Coahuila (36,829), and Sonora (33,298), which correspond to background research indicating that the most industrial activity lies in the major commercial centers of Mexico and the border states. The destinations receiving the highest total value of shipments are the Chihuahua ($3.8 billion), Distrito Federal ($3.3 billion), Tamaulipas ($2.5 billion), Mexico ($2.5 billion), and Baja California ($2.3 billion). Table 4.4 shows the destination states with the highest total value of shipments by mode and commodity. The border states mostly receive U.S. electronics exports by truck. The state of Mexico receives the highest total value of shipments of vehicles by truck. Guanajuato ranks in the top six states with the highest value of shipments by rail but not for truck. The preference for rail as the transport mode is perhaps owing to the fact that rail is accessible throughout most of the state.

4.6 CONCLUSIONS

After compiling the various data into one data set using the Statistical Analysis Software (SAS) package, exploratory analyses on the data were performed to assess the general characteristics of the data reported above. The next chapter explains the theory behind the formulation for modeling the joint mode and destination decisions. It also presents the model specification and estimation process.

Table 4.3. Origin States with Highest Total Value and Corresponding Number of Shipments by Mode

Truck				Rail			
Commodity	Origin	Total Value	Number of Shipments	Commodity	Origin	Total Value	Number of Shipments
Machinery	Texas	$2,055,928,518	70,481	Machinery	Texas	$228,722,446	3,021
	California	$946,192,463	29,609		Kentucky	$13,536,934	137
	Arizona	$427,613,645	7,165		Florida	$3,291,385	18
	Michigan	$408,897,752	5,160		Wisconsin	$3,290,060	22
	Illinois	$276,986,419	8,989		California	$2,551,745	146
	Ohio	$153,649,061	5,036		Arizona	$2,001,051	51
Electronics	Texas	$5,947,506,659	181,225	Electronics	Texas	$38,616,395	1,299
	California	$1,915,489,461	56,027		California	$4,611,043	186
	Arizona	$411,973,674	11,780		Illinois	$835,191	33
	Illinois	$134,411,947	5,473		Arizona	$689,735	72
	Michigan	$126,914,264	4,305		Pennsylvania	$634,915	20
	Missouri	$94,831,244	3,784		Missouri	$558,857	22
Vehicles	Texas	$1,346,398,868	53,639	Vehicles	Texas	$1,052,783,024	19,671
	Michigan	$1,003,850,469	9,377		Michigan	$418,495,676	6,271
	Arizona	$446,003,354	6,159		California	$16,217,808	143
	California	$187,294,056	11,505		Illinois	$811,217	19
	Tennessee	$100,520,142	4,326		Ohio	$769,315	21
	Illinois	$66,790,954	2,287		Indiana	$613,539	12

Table 4.4. Destination States with Highest Total Value and Corresponding Number of Shipments by Mode

Truck				Rail			
Commodity	Destination	Total Value	Number of Shipments	Commodity	Destination	Total Value	Number of Shipments
Machinery	Distrito Federal	$1,054,666,010	23,926	Machinery	Coahuila	$189,987,899	2,350
	Mexico	$656,209,338	17,793		Guanajuato	$27,851,058	172
	Chihuahua	$628,769,899	25,132		Distrito Federal	$19,761,827	434
	Jalisco	$591,117,471	8,230		Mexico	$9,079,866	201
	Baja California Norte	$483,460,339	17,729		Nuevo Leon	$3,267,005	101
	Sonora	$467,152,347	9,043		Sonora	$9,182,303	49
Electronics	Chihuahua	$3,108,157,703	102,492	Electronics	Coahuila	$23,961,324	798
	Tamaulipas	$1,809,217,859	39,453		Distrito Federal	$7,740,504	295
	Baja California Norte	$1,671,393,342	47,539		Tamaulipas	$4,109,082	72
	Mexico	$646,898,841	20,265		Mexico	$3,790,866	119
	Sonora	$642,210,045	17,763		Guanajuato	$3,313,166	154
	Distrito Federal	$603,889,492	20,674		Baja California Norte	$58	22
Vehicles	Mexico	$1,131,652,744	40,227	Vehicles	Distrito Federal	$816,706,616	8,527
	Distrito Federal	$826,231,212	5,605		Coahuila	$570,485,721	13,045
	Sonora	$460,170,460	6,358		Mexico	$70,972,204	2,281
	Tamaulipas	$262,148,819	8,622		Guanajuato	$55,189,864	1,913
	Baja California Norte	$146,284,505	7,907		Nuevo Leon	$5,315,013	98
	Chihuahua	$72,647,780	2,594		Puebla	$5,111,386	202

CHAPTER 5. MODEL SPECIFICATION

5.1 INTRODUCTION

By utilizing the shipment decisions revealed in U.S. Customs data of imports and exports between the U.S. and Mexico, a discrete choice model which attempts to forecast choices of shipment destination and mode of transport may be developed. This chapter presents the model structure for the analysis through a brief explanation of the theoretical framework for multidimensional choice and properties of discrete choice models.

5.2 THEORETICAL FRAMEWORK

Discrete choice models have been developed based on random utility theory. From this perspective, the alternative which is most attractive to the decision maker is chosen. Attractiveness is measured by the total utility of the alternative as compared with each of the other alternatives in the choice set that are available to the decision maker. Because the utilities are not known with certainty by the modeler, they are considered to be random variables. Utility is expressed as a sum of systematic and random components. Systematic utility results from the observed or measurable attributes affecting the choice probability. The random utility component, also interpreted as the error term, arises with the taste variations among the decision makers, attributes that cannot be quantified, deficiencies in the systematic part, and proxy variables used when data for an attribute are difficult to obtain or measure. The chosen alternative is then the one having the maximum utility of all the alternatives in a decision maker's choice set. In general, both the decision process and a joint probability distribution for the error terms are hypothesized in determining which type of discrete choice model is applicable for a particular situation.

Multinomial choice analysis deals with more than two alternatives in a choice set. Based on random utility theory, the probability of a choice being selected can be expressed in the following manner. For a decision maker n, the probability that he chooses alternative i from alternatives j in the choice set C_n is

$$P_n(i) = \Pr(U_{in} \geq U_{jn}, \forall j \in C_n)$$

With the utility function being comprised of the systematic portion denoted by V_n and a random component denoted by ε_n , the above expression becomes:

$$\begin{aligned} P_n(i) &= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n, \; j \neq i) \\ &= \Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in}, \forall j \in C_n, \; j \neq i) \end{aligned}$$

The multinomial logit model (MNL) has been a widely estimated multinomial choice model because of the ease of estimation with disaggregate data. An important assumption of logit models is that the error terms for the different choice alternatives are independently and identically Gumbel-distributed. This assumption restricts the disturbances such that they all have the same scale; in other words, they have the same variance (Ben Akiva and Lerman, 1985). In addition, the MNL model requires the independence of irrelevant alternatives (IIA), which implies that the ratio of the choice probabilities of two alternatives does not change if an alternative is added to or removed from the choice set (Ortuzar and Willumsen, 1994). This property may not be realistic under all circumstances, since it requires that the error terms be mutually independent. Mutual independence of error terms does not recognize the presence of common unobserved factors that lead to correlation of the disturbances across the alternatives, such as the combinations of modes and destinations.

5.3 DECISION STRUCTURE

5.3.1 Multidimensional Choice

In multidimensional choice, more than one decision is to be made. For example, considering that two choices, such as destination and mode, will be made, two choice sets exist, one for destination and one for mode. All the feasible combinations of these two dimensions that are available to the decision maker will comprise the multidimensional choice set for that individual. Since the elements of this choice set may share common destinations or modes, correlations among alternatives may exist. In these choice situations where IIA may be violated, the decision may be influenced by factors varying across only one dimension or across both (Ben-Akiva and Lerman, 1985). For this multiple-choice scenario of destination and mode, two types of decision structures are possible. The decisions may be made sequentially or simultaneously. A sequential decision structure assumes that a choice hierarchy exists in which one decision is made before another. A simultaneous decision structure, on the other hand, implies that the two decisions occur jointly.

In a multidimensional choice problem (such as that of mode and destination), the decision will be made according to the attributes of the mode, of the destination, and of those attributes reflecting the combination of the two. The total utility of a model can thus be expressed as:

$$U_{dm} = V_m + V_d + V_{dm} + \varepsilon_m + \varepsilon_d + \varepsilon_{dm}$$

where V_m, V_d, and V_{dm} are the systematic or measurable attributes of the total utility that vary across mode only, destination only, and both destination and mode, respectively. The ε_m, ε_d, and ε_{dm} are the random or unobserved attributes of the total utility that vary across mode only, destination only, and both destination and mode, respectively.

5.3.2 Specification of a Simultaneous Decision Structure

The decision structure hypothesized in this research is that which assumes mode and destination choices are made simultaneously. If the decision structure is actually simultaneous rather than hierarchical, no hidden attributes are shared among alternatives ($\varepsilon_d \approx 0$ and $\varepsilon_m \approx 0$).

The error term of attributes varying across both mode and destination (ε_{dm}) are then IID across the joint alternatives of mode and destination. This decision structure is modeled as a joint logit model, which may then be estimated in a number of ways.

Because the voluminous size of the BTS data requires that they be aggregated, an aggregate logit model is appropriate. The derivation of the aggregate logit model arises from the assumption of IID Gumbel-distributed, i.e., logistically distributed, error terms. The probability of a joint choice of destination and mode, represented by *i*, is given by

$$P_n(i) = \Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in}, \forall j \in C_n, j \neq i)$$

$$P_n(i) = \frac{\exp(\omega V_{in})}{\sum_j \exp(\omega V_{j'n})}$$

$$\Rightarrow \quad P_n(dm) = \frac{\exp(\omega \beta_{md} z_{md})}{\sum_{m'} \sum_{d'} \exp(\beta_{m'd'} z_{m'd'})}$$

where

V_{in} = systematic utility of choice *i* for individual *n*

β = vector of attribute parameters,

z_{md} = vector of variables for mode and destination, and

ω = scale parameter of the utilities.

The systematic utilities are assumed to be linear in parameters.

5.4 ESTIMATION OF THE AGGREGATE LOGIT MODEL

5.4.1 Maximum Likelihood Estimation

A common technique for calibrating discrete choice models is maximum likelihood estimation. A log-likelihood function describing the likelihood of an observation or choice is specified and then maximized. This method is used with samples of independent, disaggregate observations. Each observation is comprised of a vector of attributes and an indicator variable that reveals the alternative chosen. The estimates for the β's, the coefficients of the attributes, can be determined using a search process for an ascending likelihood function value. After the β's are determined, convergence is checked and the search process continues if necessary (Ben-Akiva and Lerman, 1985).

5.4.2 Berkson-Theil Estimation Method

As described in Chapter 4, the data set developed for estimation of the aggregate logit model is comprised of aggregate shipment observations. Single shipment quantities may be arbitrarily obtained by averaging attribute values. Unfortunately, disaggregating the data for this analysis does not yield very accurate estimates of the coefficients. An alternative way of estimating the β's with the aggregated data is by the Berkson-Theil method. This technique requires empirical data as well as multiple observations of a decision maker so the probabilities of each choice may be determined. In this research, the nine origins may be thought of, on the one hand, as nine different decision makers, or, on the other hand, as all decision makers in the same origin behaving the same. The estimator is derived by taking the log of the ratio of the probabilities to a probability of a reference alternative:

$$Y_n = \log\left(\frac{\Pr_n(jk)}{\Pr_n(yz)}\right) = \log\left(\frac{P_{njk}}{1-P_{njk}}\right) = \log\left(\frac{\dfrac{1}{1+\exp(\beta\,(x_{yzn}-x_{jkn}))}}{\dfrac{\exp(\beta\,(x_{yzn}-x_{jkn}))}{1+\exp(\beta\,(x_{yzn}-x_{jkn}))}}\right)$$

$$= \log\left(\frac{1}{\exp(\beta\,(x_{yzn}-x_{jkn}))}\right) = \log\big(\exp(-\beta\,(x_{yzn}-x_{jkn}))\big) = \beta\,(x_{jkn}-x_{yzn})$$

where j is the mode, either truck or rail, and k is the one of the twelve Mexican destination regions. The y and z denote the same, mode and destination, for the reference alternative. The probabilities of each choice for each origin are determined with empirical data, i.e., the number of shipments corresponding to the choice over the entire number of shipments leaving the origin. The reference alternative has been chosen as the alternative having the largest probability of occurrence. The dependent variable, the log of the ratio of probabilities, is then regressed against the difference in utilities, or attribute values to estimate β by ordinary least squares.

5.5 CONCLUSIONS

Based on previous studies of freight demand and mode choice models, a framework to estimate both destination and mode choices for U.S. exports to Mexico has been presented. This methodology will enable the estimation of modal shares of traffic between U.S. origins and Mexican destinations. The coefficients of the aggregate logit model will be calibrated using multiple regression of the attributes over the log of the probability ratio. The following chapter describes the results obtained from the aggregate logit model.

CHAPTER 6. ANALYSIS OF RESULTS

6.1 INTRODUCTION

This chapter presents the parameter estimation results of the aggregate logit model formulated in the previous chapter. This model may be used to predict the likelihood that a shipment is sent by truck or rail to a particular Mexican destination from a U.S. origin. Two approaches have been used in the calibration of the models.

6.2 RESULTS OF AGGREGATE LOGIT MODEL

The analysis of the aggregated shipment data was performed in two ways. First, the Berkson-Theil estimators were determined using linear regression separately for each of the nine origins and three commodities. The second approach was to calibrate models for each of the three commodities, with the origins pooled together.

The variables used in the regression of the log of the ratio of probabilities were attributes believed to influence mode and destination choice. As stated in the previous chapter, the data set incorporated origin and destination demographic characteristics, such as population and employment as well as the number of firms in each of the three industries — machinery, electronics, and vehicles. Another attribute is the number of shippers and warehouses within the origin and destination regions. The remaining variables are distance between the origin and destination and average value of a shipment. Note that population and employment have been scaled by a factor of 1 million, and average value per shipment has been scaled by a factor of 1 thousand. The average value of the shipment is specified as an alternative-specific variable to capture the mode likelihood effects.

The remainder of this section explains the interpretation of the coefficients of each model by the two approaches. The utility functions for the alternatives are presented for each model. Recall from the previous chapter, the specification for the aggregate logit model:

$$P_n(i) = \frac{\exp(V_{in})}{\sum_j \exp(V_{j'n})}$$

where i is the chosen alternative and j are all the choice alternatives. Recognizing that V represents the systematic utility function, V_{in} is the systematic utility function for the chosen alternative.

Using the estimated coefficients of the utility function, the systematic utility of each alternative may be determined given the attributes of the alternatives. Then, the probability that each alternative is chosen may be calculated.

6.2.1 Origin And Commodity-Specific Models

The model regression results for each origin and commodity are summarized in the tables in Appendix A, where the coefficients of the significant variables, variable t-statistics, adjusted R-squared and F-values are given. Tables 6.1, 6.2, and 6.3 show the coefficients of the utility function that yield the best models for each of the nine origins and three commodities.

Analyzing the models required several special considerations. First, the signs and magnitudes of the coefficients should correspond to prior knowledge and engineering judgment. The t-test is used to determine the significance of the coefficient. The null hypothesis of the t-test is that the coefficient is equal to zero, meaning that the variable does not affect the choice probability. The alternate hypothesis is then that the coefficient is not equal to zero, which implies that a two-tailed test will be performed. The adjusted R-squared is the coefficient of determination and indicates the percentage of the total variation in the dependent variable that is explained by the independent variables. Thus, adjusted R-squared values close to one are desirable. The F-test is used to determine the significance of the model. The null hypothesis of the F-test is that all the coefficients in the model are zero, with the alternative hypothesis being that not all of the coefficients in the model are zero. Note that in each of the model specifications with alternative-specific value coefficients, the following holds:

1. Value by truck equals the average shipment value if the chosen mode for the alternative is truck and that for the reference alternative is rail. Otherwise, value by truck is zero.
2. Value by rail equals the average shipment value if the chosen mode for the alternative is rail and that for the reference alternative is truck. Otherwise, value by rail is zero.

Analysis for Machinery (Table 6.1)

East North Central (Chicago)

$V_n = -5.29\ (1-\delta_n) - 0.002763\ (\text{DISTANCE})_n - 0.017721\ (\text{VALUE-BY-RAIL})$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (with centroid Mexico City), and zero otherwise.

The coefficients of this model are plausible insofar as the probability of the mode and destination choice will decrease with longer distances. Average value is defined as an alternative-specific variable for the rail mode. This specification captures only the effect of value on the mode choice, not on destination choice, and allows each mode to have differing effects on the utility. All the coefficients are significant at the 5-percent level. For every 161-km (100-mile) increase in transport distance, the utility of the alternative decreases by about 28 percent. A $1 million increase in value of a shipment going by rail causes a 1.8-percent decrease in utility, which may be interpreted as a preference for the truck mode. These variables explain 45 percent of the variation in the dependent variable, the log of the probability ratio. The model is significant at the 1-percent level, since the F-test shows that there is a 0.0032 probability that all the variables are not equal to 0 by chance.

Table 6.1. Aggregate Logit Model Estimation Results for Machinery Shipments from Nine Origins

Parameter	*Chicago*	*Birmingham*	*Philadelphia*	*Salt Lake City*	*Boston*	*Atlanta*	*Sioux City*	*Dallas*	*San Francisco*
INTERCEPT*	-4.29	-1.14	4.037	-2.969	15.144	0.9471	-3.46995	-3.0147	-2.7462
	-9.44	**-0.533**	**2.97**	**-2.7**	**2.5**	**0.841**	**-7.43**	**-4.591**	**-2.666**
DISTANCE	-0.002763	-0.00187	-0.003319	-0.001865		-0.0046		-0.0056	-0.00197
	-2.74	**-1.507**	**-5**	**-2.363**		**-8.675**		**-5.265**	**-3.039**
EMPLOYMENT						0.67979			
						4.287			
POPULATION		0.1504		0.090757				0.1234	0.208702
		1.49		**1.543**				**1.966**	**3.791**
SHIPPERS			0.0564		0.1387				
			5.892		**3.129**				
VALUE-BY-RAIL	-0.017721		-0.2133	-0.057837	-0.3596	-0.0267	-0.1387		-0.17034
	-2.42		**-7.82**	**-1.498**	**-3.437**	**-6.132**	**-3.482**		**-4.624**
R-square	0.513	0.1712	0.8621	0.3738	0.5319	0.882	0.4311	0.5974	0.6888
Adj R-sq	**0.4521**	**0.0607**	**0.8303**	**0.203**	**0.4468**	**0.8607**	**0.3955**	**0.5572**	**0.6338**
F-Value	8.427	1.549	27.091	2.189	6.249	36.001	12.125	14.84	12.54
F Prob	0.0032	0.2446	0.0001	0.1469	0.0154	0.0001	0.0031	0.0001	0.0001

* The intercept is included in the utilities of all alternatives except the reference alternative.

Note: These are the coefficients of the utility function of an alternative which may be used to determine the probability that an alternative is chosen. The alternative-specific value constants take on values for the corresponding mode utilities and zero otherwise.

East South Central (Birmingham)

V_n = -1.14 (1-δ_n) - 0.00187 (DISTANCE)$_n$ +0.1504 (POPULATION)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

For this model having an adjusted R-squared of 0.061, the coefficients are only significant at the 20-percent level. The variables, distance and population, do not sufficiently describe the utility function. Also, the likelihood that all the coefficients are zero by chance is high, at 50 percent.

Mid-Atlantic (Philadelphia)

V_n = 4.037 (1-δ_n) - 0.003319 (DISTANCE)$_n$ + 0.0564 (SHIPPERS)$_n$ -0.2133 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

The best model calibrated for Mid-Atlantic origin had an adjusted R-squared of 0.83 and included the variables distance, number of shippers, and value-by-rail. The coefficient of distance (-0.003319) indicates that an alternative with a destination 161 km (100 miles) further than another destination will have a 33-percent decrease in utility. For an increase of shipment value by $1000 for a good transported by rail, the utility of the alternative being chosen decreases by about 21 percent. Thus, from the Mid-Atlantic region, shipments are more likely to go by truck than by rail. The number of shippers in a region also influences the probability that an alternative is chosen. The higher the number located in the destination region, the higher the utility of that region.

Mountain (Salt Lake City)

V_n = -2.969 (1-δ_n) - 0.001865 (DISTANCE)$_n$ + 0.090757 (POPULATION)$_n$ -0.05784 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Sonora/Sinaloa Region (Hermosillo), and zero otherwise.

For shipments originating in the Mountain region, the variables that contribute to the likelihood of an alternative being chosen are distance, population, and value. Distance is significant at the 5-percent level and an increase in 161 km (100 miles) of transport distance will cause about a 19-percent decrease in utility. A higher shipment value will decrease the utilities with the rail modal alternative. A 9-percent increase in utility will result from an increase in population by 1 million. These variables explain 20 percent of the log of the probability ratios. The model is significant at the 30-percent level.

New England (Boston)

$V_n = 15.144\ (1-\delta_n) + 0.13865\ (\text{SHIPPERS})_n - 0.35964\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

In this model, value and the number of shippers involved in transport of machinery and mechanical appliances captured 45 percent of the variation in the log of the probability ratios. The coefficient of value by rail is negative and the largest in magnitude of all the models, indicating that shipments from the New England region are more likely to be shipped by truck. This model is significant at the 5-percent level.

South Atlantic (Atlanta)

$V_n = 0.9471\ (1-\delta_n) - 0.0046\ (\text{DISTANCE})_n + 0.68\ (\text{EMPLOYMENT})_n - 0.0267\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

Higher employment in a destination region should cause an increase in the utility of that alternative. The likelihood of the alternative being chosen would be greater, since more of the population have disposable income to purchase goods. All the coefficients are significant at the 5-percent level, and the variables explain 86 percent of the variation of the dependent variable. An increase in distance by 161 km (100 miles) will cause a 46 percent decrease in the utility of an alternative. If the value of the shipments increase by $1000, the utility of each alternative will also decrease by almost 3 percent if it is going by rail.

West North Central (Sioux City)

$V_n = -3.47\ (1-\delta_n) - 0.1387\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

These model coefficients are significant at the 5-percent level. The adjusted R-squared value is 0.4. An increase in value of the shipments by $1000 will cause the utility of the alternatives to decrease by approximately 14 percent if the shipment goes by rail. Longer distances have a diminishing effect on probability. Since the model's F-value is 12.125, the null hypothesis that all coefficients in the model are zero may be rejected at the 5-percent level.

West South Central (Dallas)

V_n =-3.0147 $(1-\delta_n)$ - 0.00557 (DISTANCE)$_n$ + 0.12339 (POPULATION)$_n$
where δ_n is a binary variable equal to one for the reference alternative, the Chihuahua/Durango Region (Chihuahua), and zero otherwise.

For shipments leaving this region, all coefficients are significant at the 10-percent level. The magnitude of the distance coefficient compared with the other origins reveals that small differences in transport distance have a greater effect on probability than for shipments originating from other regions. An increase in population by 1 million people will cause the utility to increase by about 12 percent. These variables explain 56 percent of the total variation of the log of the probability ratio.

West (San Francisco)

V_n = -2.7462 $(1-\delta_n)$ - 0.00197 (DISTANCE)$_n$ + 0.208702 (POPULATION)$_n$ -0.17034 (VALUE-BY-RAIL)$_n$
where δ_n is a binary variable equal to one for the reference alternative, the Baja California Region (Tijuana), and zero otherwise.

In this model, distance, value, and population are significant at the 5-percent level. The variables are able to explain 63 percent of the total variation of the log of the ratio of probabilities. The value for shipments by rail is again negative, indicating the greater likelihood that rail is the chosen mode. A population increase of a million will cause a 21-percent increase in utility, whereas, a distance increase of 161 km (100 miles) will cause about a 20-percent decrease in the utility of the alternative.

Analysis for Electronics (Table 6.2)

East North Central (Chicago)

V_n = 1.885 $(1-\delta_n)$ - 0.002742 (DISTANCE)$_n$ + 0.090661 (FIRMS)$_n$ - 0.33556 (VALUE-BY-RAIL)$_n$
where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

Table 6.2. Aggregate Logit Model Estimation Results for Electronics Shipments from Nine Origins

Parameter	*Chicago*	*Birmingham*	*Philadelphia*	*Salt Lake City*	*Boston*	*Atlanta*	*Sioux City*	*Dallas*	*San Francisco*
INTERCEPT	1.884625	-3.231444	-3.296354	-3.2575	-4.1022	19.1344	-1.59615	-4.095	-3.6766
	1.078	**-6.219**	**-7.793**	**-3.572**	**-8.398**	**4.716**	**-1.825**	**-5.18**	**-4.414**
DISTANCE	-0.002742	-0.002614	-0.00421	-0.002778	-0.0043		-0.00304	-0.0042	-0.00233
	-3	**-2.997**	**-3.12**	**-3.229**	**-3.107**		**-2.672**	**-2.673**	**-4.432**
POPULATION				0.205205					0.182074
				0.3136					**3.957**
FIRMS	0.090661					0.36781			
	3.376					**5.655**			
SHIPPERS								0.0208	
								1.743	
VALUE-BY-RAIL	-0.335562	-0.084175	-0.056445	-0.248954		-0.5032	-0.29207	-0.0604	-0.0587
	-6.493	**-2.694**	**-3.308**	**-2.939**		**-6.269**	**-3.216**	**-1.979**	**-3.792**
VALUE PER MILE (RL)						-57.68			
						-4.86			
R-square	0.7609	0.5128	0.6717	0.7009	0.4458	0.7973	0.5366	0.4828	0.7313
Adj R-sq	**0.7096**	**0.4378**	**0.612**	**0.5888**	**0.3996**	**0.7365**	**0.4594**	**0.372**	**0.6737**
F-Value	14.848	6.841	11.253	6.25	9.652	13.11	6.948	4.356	12.698
F Prob	0.0001	0.0093	0.0022	0.0172	0.0091	0.0008	0.0099	0.023	0.0003

* The intercept is included in the utilities of all alternatives except the reference alternative.

Note: These are the coefficients of the utility function of an alternative which may be used to determine the probability that an alternative is chosen. The alternative-specific value constants take on values for the corresponding mode utilities and zero otherwise.

Seventy percent of the variation in the log of the probability ratio is explained by distance, number of firms, and value by the mode choice. For every additional firm in the electronics sector in the Mexican destination, the alternative's utility increases by 9 percent. Truck shipments are preferred to rail shipments in terms of value, since the coefficient for value by rail is negative.

East South Central (Birmingham)

$V_n = -3.23\ (1-\delta_n)\ -0.002614\ (\text{DISTANCE})_n\ -0.084175\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

All coefficients in this model are significant at the 5-percent level. Distance and value explain the probability of the alternative being selected. An additional 161 km (100 miles) in transport distance will cause the utility of the mode-destination alternative to decrease by approximately 26 percent. An increase of $1000 in shipment value causes about a 8.4-percent decrease in utility of the alternative if rail is used to transport the goods.

Mid-Atlantic (Philadelphia)

$V_n = -3.2964\ (1-\delta_n)\ -\ 0.00421\ (\text{DISTANCE})_n\ -0.056445\ (\text{VALUE-BY-RAIL})$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

The adjusted R-squared for this model is 0.612. Distance and value by rail are again both significant in determining the likelihood of a mode-destination choice. They are also both negative, indicating that truck is the preferred mode of transport, since rail distances tend to be less direct than highway distances.

Mountain (Salt Lake City)

$V_n = -3.2575\ (1-\delta_n)\ -\ 0.002778\ (\text{DISTANCE})_n\ +\ 0.2052\ (\text{POPULATION})_n\ -\ 0.249\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Sonora/Sinaloa Region (Hermosillo), and zero otherwise.

For electronic shipments originating in the Mountain region, most shipments travel to the Sonora/Sinaloa Region, which is close to the origin. The variables that contribute to the likelihood of an alternative being chosen are distance, population, and value for rail shipments. All the coefficients are significant at the 5-percent level. Truck shipments are more likely to occur, since the utility of an alternative is negatively influenced for rail alternatives.

New England (Boston)

$V_n = -4.1\ (1-\delta_n)\ -0.0043\ (\text{DISTANCE})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

For shipments originating in New England, the destination and mode likelihoods are explained by transport distance. Every additional 161 km (100 miles) in length of haul to the destination region causes a 4.3-percent decrease in the utility of the alternative. All the coefficients are significant at the 5-percent level. Distance is able to capture 40 percent of the variation in the log of the probability ratio.

South Atlantic (Atlanta)

$V_n = 19.134\ (1-\delta_n) + 0.36781\ (\text{FIRMS})_n\ -0.5032\ (\text{VALUE-BY-RAIL})_n\ -57.68\ (\text{VALUE PER MILE -RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

All the coefficients are significant at the 5-percent level, and the variables explain the variation in the log of the probability ratio 74 percent of the time. An additional firm in the electronics sector in the destination region causes an increase in utility by 37 percent. Shipments are more likely to be transported by truck than by rail. The higher the shipment value per mile to the destination region, the less likely the shipment will be transported by rail.

West North Central (Sioux City)

$V_n = -1.596\ (1-\delta_n) - 0.00304\ (\text{DISTANCE})_n\ -0.2921\ (\text{VALUE-BY-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Nuevo Leon Region (Monterrey), and zero otherwise.

These model coefficients are significant at the 5-percent level. The adjusted R-squared value is 0.46 for the two variables. Every 161 km (100 miles) of additional transport distance lowers the utility of the alternative by 30 percent. An increase in value of the shipment by $1000 will cause the utility to decrease by approximately 29 percent if the shipment is transported by rail. Thus, truck shipments are favored. The model is significant at the 1-percent level.

West South Central (Dallas)

V_n = -4.095 (1-δ_n) - 0.00422 (DISTANCE)$_n$ -0.06043 (VALUE-BY-RAIL)$_n$ + 0.0208 (SHIPPERS)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Chihuahua/Durango Region (Chihuahua), and zero otherwise.

The destination of shipments of electronic components, electrical equipment, and the like has a substantial impact on choice for shipments from the Dallas region. The coefficients are significant at approximately the 10-percent level. The location of shippers and warehouses in the destination regions seems to have an impact on whether a shipment goes to that destination. The significance of this attribute in the model may perhaps suggest that electronics shipments from Dallas tend to be distributed to the destination population. Value of the shipments by rail decreases utility by 6 percent for rail shipments.

West (San Francisco)

V_n = -3.677 (1-δ_n) - 0.00233 (DISTANCE)$_n$ + 0.18207 (POPULATION)$_n$ -0.0587 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Baja California Region (Tijuana), and zero otherwise.

Distance, population, and shipment value by rail influence the probability that the alternative is chosen. Shipments are more likely to be transported by truck, but not more so than electronics shipments originating in Chicago, Atlanta, or Sioux City. The variables are able to explain 67 percent of the total variation in the log of the probability ratio. The model's F-value is 12.698, which indicates rejection of the null hypothesis that all coefficients are zero at the 1-percent level.

Analysis for Vehicles (Table 6.3)

East North Central (Chicago)

V_n = 3.795 (1-δ_n) - 0.00286 (DISTANCE)$_n$ + 0.05876 (FIRM-SHIP-WARE)$_n$ + 0.01714 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

Table 6.3. Aggregate Logit Model Estimation Results for Vehicle Shipments from Nine Origins

Parameter	*Chicago*	*Birmingham*	*Philadelphia*	*Salt Lake City*	*Boston*	*Atlanta*	*Sioux City*	*Dallas*	*San Francisco*
INTERCEPT	3.794784	-4.680686	1.58582	-5.256571	1.8532	0.09553	6.42797	-5.7867	-4.1208
	1.839	**-13.369**	**0.733**	**-7.223**	**0.909**	**0.046**	**2.15**	**-9.338**	**-3.701**
DISTANCE	-0.00286		-0.004375	-0.00121	-0.0056	-0.003	-0.00187	-0.0041	-0.0021
	-3.374		**-2.36**	**-2.261**	**-1.985**	**-2.546**	**-1.503**	**-2.27**	**-2.599**
EMPLOYMENT					1.113				0.596874
					2.61				**2.907**
POPULATION			0.246391			0.18874			
			2.507			**2.12**			
FIRMS		0.212515					0.47332		
		4.256					**3.431**		
FIRM-SHIP-WARE	0.058758								
	4.642								
VALUE									
VALUE PER MILE (RL)		-76.116108							
		-2.904							
VALUE-BY-RAIL	-0.060537		-0.415281		-0.3876	-0.0603	-0.23695		
	-2.645		**-2.746**		**-2.716**	**-1.958**	**-3.323**		
R-square	0.7153	0.6717	0.5831	0.3173	0.8643	0.5676	0.6268	0.2326	0.4453
Adj R-sq	**0.6496**	**0.5987**	**0.4267**	**0.2552**	**0.6609**	**0.4497**	**0.5025**	**0.1875**	**0.366**
F-Value	10.886	9.206	3.729	5.112	4.248	4.813	5.04	5.153	5.619
F Prob	0.007	0.0067	0.0606	0.045	0.1964	0.0223	0.0255	0.0365	0.0162

* The intercept is included in the utilities of all alternatives except the reference alternative.

Note: These are the coefficients of the utility function of an alternative which may be used to determine the probability that an alternative is chosen. The alternative-specific value constants take on values for the corresponding mode utilities and zero otherwise.

This model is significant at the 1-percent level. The coefficients of distance, number of firms, shippers, and warehouses in the destination region, and shipment value are significant at the 5-percent level and therefore influence the probability of the alternative being chosen. A 161-km (100-mile) increase in transport distance causes a 3-percent decrease in utility and thus decreases the probability of the alternative being chosen. A $1000 increase in shipment value will cause about a 6-percent decrease in utility if the mode is rail. As a result, shipments are more likely to be transported by truck in terms of value.

East South Central (Birmingham)

$V_n = -4.681\ (1-\delta_n) + 0.212515\ (\text{FIRMS})_n - 76.12(\text{VALUE PER MILE-RAIL})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

The number of firms and the shipment value per mile transported influence the likelihood of an alternative being chosen, since it is influenced by the magnitude of the utility of the alternative. An additional firm in the destination region in the automobile industry increases the utility by 21 percent. High valued shipments being transported short distances are more likely to go by truck. The adjusted R-squared is 0.60. The coefficients are significant at the 5-percent level.

Mid-Atlantic (Philadelphia)

$V_n = 1.58582\ (1-\delta_n) - 0.004375\ (\text{DISTANCE})_n - 0.4153\ (\text{VALUE-BY-RAIL})_n + 0.2464\ (\text{POPULATION})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

This model results in an adjusted R-squared of 0.427 calibrated for the vehicle commodity group. The model is significant at the 10-percent level, with coefficients being significant at the 5-percent level. Truck shipments are highly preferred in terms of shipment value, since the magnitude of the value by rail coefficient is large and negative. A million-person increase in population increases the utility by 25 percent. The likelihood of a mode/destination alternative would be greater for alternatives having a shorter length of haul.

Mountain (Salt Lake City)

$V_n = -5.26\ (1-\delta_n) - 0.00121\ (\text{DISTANCE})_n$

where δ_n is a binary variable equal to one for the reference alternative, the Sonora/Sinaloa Region (Hermosillo), and zero otherwise.

Distance contributes to probability in shipments from the Mountain region but to a smaller degree than for other origins (except Birmingham, which is unaffected by distance). For a 161-km (100-mile) increase in length of haul, the utility of the alternative decreases by about 1 percent. Distance is capable of explaining only 26 percent of the variation in the log of the ratio of the probabilities. The model is significant at the 5-percent level.

New England (Boston)

V_n = 1.8532 $(1-\delta_n)$ -0.00561 (DISTANCE)$_n$ + 1.113 (EMPLOYMENT)$_n$ -0.38762(VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

For shipments originating in the Boston area, distance, employment, and value for rail shipments affect the likelihood that an alternative is chosen. The employment coefficient indicates that the higher the employment, the greater the utility of the alternative. Shipments may be more likely to go to a destination having higher employment, since higher employment means greater disposable income. This result seems to correspond to a priori expectations that higher purchasing power on finished products causes a higher demand for the good. Since rail shipment values decrease the utility, truck seems to be the preferred mode in terms of value of the shipment. Alternatives with longer distances also have lower utilities.

South Atlantic (Atlanta)

V_n = 0.09553 $(1-\delta_n)$ - 0.003 (DISTANCE)$_n$ + 0.18874 (POPULATION)$_n$ - 0.0603 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

Distance, population, and shipment value by rail are significant in capturing the influential factors of shipments moving from Atlanta. An additional $1000 in shipment value decreases the alternative's utility by 6 percent if the mode is rail. Utility decreases by 3 percent for an increase in length of haul of 161 km (100 miles). The demand for goods is reflected in the population variable, since a 1 million-person increase in population causes an increase in utility by 19 percent. The adjusted R-squared for this model is 0.45; all coefficients are significant at the 10-percent level.

West North Central (Sioux City)

V_n = 6.43 (1-δ_n) -0.00187 (DISTANCE)$_n$ + 0.47332 (FIRMS)$_n$ -0.2367 (VALUE-BY-RAIL)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

The variables distance, number of firms in the automobile industry in the destination region, and shipment value by rail are the attributes found to capture the utility of the alternative. They are significant at the 5-percent level except distance, which is significant at the 20-percent level, and the adjusted R-squared is 0.50. A $1000 increase in shipment value causes the utility of the rail alternatives to decrease by 24 percent. An additional firm increases the utility by 47 percent.

West South Central (Dallas)

V_n = -5.79 (1-δ_n) - 0.00406 (DISTANCE)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Central Region (Mexico City), and zero otherwise.

Only distance seems to influence probability for shipments from the Dallas region. Its coefficients are significant at the 5-percent level; however, the resulting adjusted R-squared is low at 0.19. Not all the factors influencing shipment mode and destination choice are accounted for by this model.

West (San Francisco)

V_n = -4.1208 (1-δ_n) - 0.0021 (DISTANCE)$_n$ + 0.5969 (EMPLOYMENT)$_n$

where δ_n is a binary variable equal to one for the reference alternative, the Baja California Region (Tijuana), and zero otherwise.

Distance and employment are the variables that influence the probability that a shipment is transported by truck or rail to a destination region. These coefficients are significant at the 5-percent level. The adjusted R-squared value is 0.366.

6.2.2 Pooled Origins, Commodity-Specific Models

While distance and value were consistently explanatory variables for probability, the models for each commodity with pooled origins include the number of firms, the number of shippers and warehouses, and the combined number of firms, shippers, and warehouses to characterize the destination. Table 6.4 illustrates the components of each of the commodity-specific models for the pooled origins.

Table 6.4 Aggregate Logit Model Estimation Results for Pooled Origins by Commodity Group

Parameter	Machinery SCH_B=84	Electronics Sch_B=85	Vehicles Sch_B=87
INTERCEPT	-3.099 **-1.39**	-3.078291 **-3.981**	5.145564 **1.899**
DISTANCE	-0.002644 **-3.54**	-0.002409 **-2.608**	-0.003199 **-4.216**
FIRMS	0.189871 **2.006**		0.482133 **3.902**
SHIPPERS	-0.10561 **-1.685**		
FIRM-SHIP-WARE		0.017619 **2.052**	
VALUE-BY-RAIL	-0.035396 **-2.617**	-1.01438 **-3.003**	-0.064839 **-2.405**
R-square	0.5478	0.4833	0.6018
Adj R-square	**0.4473**	**0.3864**	**0.5271**
F-value	5.451	4.989	8.06
F Prob	0.0047	0.0125	0.0017

The intercept is included in the utilities of all alternatives except the reference alternative. NOTE: These are the coefficients of the utility function of an alternative that may be used to determine the probability that an alternative is chosen. The alternative-specific value constants take on values for the corresponding mode utilities and zero otherwise.

For machinery shipments, distance, shipment value for rail, the number of shippers and warehouses, and the number of firms in the destination region are capable of modeling the joint mode and destination decisions. The model for electronic shipments consists of the distance, shipment value for rail, and the combined number of firms, shippers, and warehouses. Finally, for vehicle shipments, distance, shipment value for rail, and the number of firms are found to influence the choice probability. Longer distances impact utility more so for automobile shipments than for machinery and electronics. The shipment value by rail affects the utility the greatest for electronics shipments, perhaps because electronics are a more valuable commodity and have a higher risk of damage during shipping. Therefore, the truck mode would be preferred over rail. The adjusted R-squared values of the models are comparable to the origin-commodity specific models. The variables are significant primarily at the 5-percent and 10-percent levels, which suggests that pooling the data results in satisfactory models overall. However, since aggregation causes differences within the data to be averaged out, the predictions often may not be accurate. Each of the three models are significant at the 1-percent level.

6.3 SUMMARY OF RESULTS

For shipments of machinery and mechanical appliances, almost all the models except for Sioux City and Boston include distance as a factor contributing to selecting a mode and destination alternative. Employment plays a role in the probability of choosing an alternative for shipments originating in the Atlanta region, whereas population affects shipments from Birmingham, Salt Lake City, Dallas, and San Francisco. The value of the shipment to the destination by rail was significant in all cases but Birmingham and Dallas. The number of shippers and warehouses was a significant variable in determining the probability for Philadelphia and Boston. The number of firms involved in machinery and mechanical appliances, on the other hand, did not influence probability for any origin region. The degree of variation that was able to be explained by these variables changed from model to model. Significant models for Birmingham and Salt Lake City could not be estimated with the variables in the data set. Table 6.5 compares these calculations for the predicted probabilities for a machinery shipment originating in Chicago. The origin-specific model is a better predictor for these observations than the pooled model for machinery shipments.

For shipments of electronics and electrical equipment, distance had the greatest impact on probability for shipments originating in Philadelphia, Boston, and the Dallas region. The number of shippers and warehouses in the Mexican destination region seemed to have an effect on determining probability for shipments from the Dallas region. Value-per-mile shipped by rail also had a role in affecting the likelihood of an alternative being chosen for shipments from Atlanta. High-value shipments of electronics and electrical equipment traveling short distances are more likely to go by truck.

The models for vehicle shipments had adjusted R-squared values that lie just around the 0.5 mark for the most part. Nearly all the models included a destination attribute; however, the shipment value for rail entered into about half the utilities. Shipment value does not influence mode choice for Birmingham, Salt Lake City, Dallas, or San Francisco. Truck shipments are slightly preferred for Chicago and Atlanta The number of firms, shippers, and warehouses was significant for vehicle shipments from Chicago, indicating that this region may be an exporter of automobile parts.

Table 6.5. Sample Probability Calculations for Machinery Shipments from San Francisco Using the Origin-Commodity Specific Aggregate Logit Model

Mode	Destination Region	Distance	Population	Shipment Value (Rail)	Utility	exp(utility)	Probability	Number of Shipments
Truck	Baja California	737.984	3957238	0	-3.37415	0.03425	0.09770	3014
Rail	Baja California	606	3957238	$17,246	-6.05178	0.00235	0.00671	207
Truck	Chihuahua/Durango	1503.57	3791251	0	-4.91699	0.00732	0.02088	644
Rail	Chihuahua/Durango	1507	3791251	$17,246	-7.86139	0.00039	0.00110	34
Truck	Coahuila	1701.78	1972340	0	-5.68708	0.00339	0.00967	298
Rail	Coahuila	2022.5	1972340	$17,246	-9.25654	0.00010	0.00027	8
Truck	Nuevo Leon	1984.17	3098736	0	-6.00830	0.00246	0.00701	216
Rail	Nuevo Leon	1745	3098736	$17,246	-8.47478	0.00021	0.00060	18
Truck	South Gulf	2536.7	3893146	0	-6.93099	0.00098	0.00279	86
Rail	South Gulf	3803	3893146	$17,246	-12.36324	0.00000	0.00001	0
Truck	South Pacific	1981.2	8850693	0	-4.80201	0.00821	0.02343	723
Rail	South Pacific	2296	8850693	$17,246	-8.35980	0.00023	0.00067	21
Truck	Sonora/Sinaloa	1140.7	4027660	0	-4.15279	0.01572	0.04484	1383
Rail	Sonora/Sinoloa	1113.36	4027660	$17,246	-7.03657	0.00088	0.00251	77
Truck	Tamaulipas	1982.64	2249581	0	-6.18251	0.00207	0.00589	182
Rail	Tamaulipas	2220	2249581	$17,246	-9.58775	0.00007	0.00020	6
Truck	Veracruz	2208.08	6228239	0	-5.79627	0.00304	0.00867	267
Rail	Veracruz	2589	6228239	$17,246	-9.48433	0.00008	0.00022	7
Truck	Central	2215.84	27073577	0	-1.46109	0.23198	0.66176	20414
Rail	Central	2179.5	27073577	$17,246	-4.32715	0.01321	0.03767	1162
Truck	North Central	1871.2	7981762	0	-4.76665	0.00851	0.02427	749
Rail	North Cental	1643.4	7981762	$17,246	-7.25553	0.00071	0.00201	62
Truck	West	1842.56	10104041	0	-4.26731	0.01402	0.03999	1234
Rail	West	2165.22	10104041	$17,246	-7.84060	0.00039	0.00112	35

6.4 CONCLUSIONS

Concerns regarding the high volumes of surface traffic crossing the U.S.-Mexico border have prompted the collection of new freight data on a national level. These data were used to estimate the aggregate logit models for nine U.S. origins. Coefficients of the utility functions can be applied to predict the likelihoods of shipment transport mode and destinations. Distance has been used to capture the level of service attributes of each mode, while population, employment, and the number of firms in the industry reflect the demand for machinery and mechanical appliances in the Mexican destination region. The value of the shipments by rail is a major factor in capturing the mode likelihood. Also, the ability of these variables to predict the probabilities depends on the commodity. In terms of the two approaches to estimation, the origin and commodity-specific models are more accurate than the models with pooled origin because the differences among the data are averaged out in the calibration process. For predictive purposes, the origin and commodity-specific models should be used.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

In this study, aggregate logit models for machinery, electronics, and vehicle commodity groups were estimated to model mode and destination choice jointly. The analysis of U.S. export shipment data was conducted for U.S. and Mexican regions. Models were estimated for each commodity, first for each of the nine U.S. origin regions, and then for the entire U.S. with the origins pooled. The data set was comprised primarily of export data of aggregated shipments obtained from the BTS. Supplementary data, such as distance and employment, were added to provide destination and modal attributes to determine the factors that influence the probability of a mode/destination choice.

Given the available data, most of the 27 aggregate logit models estimated for each commodity and origin can be applied to predict proportions of traffic traveling by truck or rail to one of 12 Mexican destinations. Several models were not significant at the 10-percent level. None of the data set variables were sufficient in capturing the influential factors in mode/destination choice. The results of the model estimation confirm the conclusions observed in the literature review of past modeling efforts that mode, commodity, and market attributes are necessary in modeling freight decision-making behavior. Distance and market attributes are necessary to adequately capture influences on mode and destination choice. Average shipment value was a significant shipment characteristic for rail in predicting the likelihood of an alternative being chosen. In many cases, such market attributes as employment were significant variables in capturing mode-destination choice. The coefficient estimates using ordinary least squares were of the correct sign and significant. However, implementation of these models requires additional study in determining the accuracy of the predictive ability, perhaps by testing the models using recently released data.

With the origin-commodity specific aggregate logit models, a Poisson regression model could be calibrated for the frequency of shipments that would enable estimating the number of shipments going by each mode-destination alternative. If the number of shipments generated at each origin can be estimated, then interregional flows can be determined. For a more disaggregated analysis, models can also be specified for specific U.S. and Mexican states.

There are some limitations to this work. Variables that have been found to affect mode choice and demand could not be included because the data were unavailable. Provided these data can be obtained, other variables that may be used in future work include transport system characteristics at both the origin and destination (e.g., mode accessibility). This type of attribute is particularly important for rail. Another attribute that could be included in the model specification is the cost for transport of the shipment by each mode. The more economical the rate, the more likely the mode will be chosen for transport. However, commodity attributes (e.g., probability of loss and damage by mode, seasonality, and whether the commodity is a bulk good) are also factors and should be included if data are available.

In addition to the limited variables available for model calibration, another limitation is the nature of the shipment data used in this analysis. Individual shipments have been aggregated by month, origin, destination, and mode because of the voluminous number of export shipments. If disaggregated data become available, a discrete choice model, such as the nested logit model, may be calibrated to model individual decision-making behavior. Furthermore, the aggregated shipment observations used in the estimation indicated only the mode of shipment across the border, and not whether transfers from other modes had been made (nor where they were made). Most likely shipments are not transported by one mode for the entire trip. Thus, the transport route would be of interest.

Further work may also try to include more than two modes in the analysis. Since maritime transport is often available to the decision maker, it may be desired as a modal alternative in his/her choice set. An effort to better capture the shifts of trade flows by mode was made by attempting to include waterborne commerce in this data set. The Waterway CD-ROM, a compilation of data from the Army Corps of Engineers Navigation Data Center, Bureau of Census, Coast Guard, and Vanderbilt University/Oak Ridge National Laboratory, is also distributed by the BTS. Of particular interest are the National Waterway Network, Waterborne Commerce Statistics Center (WCSC) files and the U.S. Bureau of the Census files. The National Waterway Network provides information on U.S. ports and waterways that may be used to build a GIS layer. The WCSC files document the following:

- domestic and foreign tonnages for U.S. ports
- state-to-state tonnages for major commodity groups
- state summary of commodity movements
- tonnages from states to domestic, foreign, and to itself
- National Waterway Network link commodity data
- summary of all cargo and loaded barge movements

The U.S. Bureau of the Census files provide information on vessel entrances and clearances. No commodity information is revealed for the movement, and only the next country of port of call for the vessel is designated if the destination is foreign. However, the cargo import and export files specify whether the cargo is bonded or released from bond in the U.S., or in transit to another country, U.S. port of origin, foreign port of destination, 2-digit code for commodities, and cargo weight. The U.S. port or waterway code of origin is defined according to that used by the Army Corps of Engineers. The foreign port of destination is the U.S. Bureau of Census 5-digit Schedule K code. Twenty-nine different codes exist for Mexican ports. Note again that true origins and destinations are not indicated in these files. The 2-digit commodity code is defined by the U.S. Army Corps of Engineers, and weight is specified in short tons.

Although a more complete model would include maritime transport for better forecasting of modal shifts, this inclusion is not possible at this time owing to resources available and data issues. For example, the three commodities of interest (i.e., electronics, machinery, and vehicles) are aggregated into the same category without possibility for distinction. Thus, including waterborne shipments in the data set was not feasible. Furthermore, waterway locations often transcend state borders as defined by the Army Corps of Engineers and would thus also require disaggregation in order to determine the shipment's state of origin. Despite its growing role in intermodal transportation, its transport role is still limited and not competitive with truck or rail in terms of time, cost, and service. Use of this mode is primarily for transporting bulk or unique goods that cannot be easily sent by another mode (Fedorowicz, 1994). Thus, for machinery, electronics, and vehicles, the use of maritime transport is minimal. However, long-term strategic planning models should include waterway and intermodal transport options. Furthermore, these models should preferably be calibrated using shipment data collected and maintained by one agency in order to most accurately reflect freight movement.

REFERENCES

Association of American Railroads (AAR). 1992. *The Intermodal Competition Model: User's Manual.* Cambridge: A&L Associates.

Bancomext Trade Directory of Mexico. 1993. National Foreign Trade Bank. Mexico, D.F.: Bancomext.

Ben-Akiva, Moshe, and Stephen R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge: MIT Press.

Bureau of Transportation Statistics (BTS). 1993-96. *National Transportation Atlas Databases,* CD-ROM, U.S. Department of Transportation, Bureau of Transportation Statistics.

Bureau of Transportation Statistics (BTS). January 1995. *Product Availability Announcement.* Washington, D.C.: U.S. Department of Transportation (U.S. DOT).

Bureau of Transportation Statistics (BTS). January 1995. DOT/Census Transborder Surface Trade Flow Data: U.S./Mexico; U.S./Canada Data Documentation for April 1994 through September 1994 Data Summary, Readme File, Surface Transborder Commodity Data Diskettes. Washington, D.C.: U.S. DOT.

Bureau of Transportation Statistics (BTS). April 1995. DOT/Census Transborder Surface Trade Flow Data: U.S./Mexico; U.S./Canada Data Documentation starting April 1995 Summary, Readme File, Surface Transborder Commodity Data Diskettes. Washington, D.C.: U.S. DOT.

Characteristics and Changes in Freight Transportation Demand. 1993. NCHRP Project 8-30 Interim Report Draft (August).

Chiang, Yu-Sheng. 1979. A Policy Sensitive Model of Freight Demand. Doctoral dissertation, Massachusetts Institute of Technology (MIT).

Clement, Norris C., Stephen R. Jenner, Paul Ganster, and Andrea Setran. 1989. *Maquiladora Resource Guide: Exploring the Maquiladora/In-Bond Option in Baja California, Mexico.* San Diego: Institute for Regional Studies of the Californias.

County Business Patterns. 1996. Office of Domestic Commerce, U.S. Department of Commerce.

Courtney, Alan C. 1995. "Entering the Mexico Market: A Transportation Assessment." *TR News* 179 (July-August): 9-14, 45.

Dennis, Scott M. 1989. *The Cross Elasticity Model.* AAR.

Fedorowicz, Jan, Marvin Bedward, JoAnna Townsend, and Ruth Rayman. 1994. *Business Mexico.* Chicago: Probus Publishing Company.

Fiscal Notes. October 1995. John Sharp, Texas Comptroller of Public Accounts. Austin: Comptroller of Public Accounts.

Guélat, Jaques, Michael Florian, and Teodor Gabriel Crainic. February 1990. "A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows." *Transportation Science*, 24, no. 1:26.

Harker, Patrick T. 1987. *Predicting Intercity Freight Flows.* Netherlands: VNU Science Press.

Hartwig, James, and William Linton. 1974. *Disaggregate Modal Choice Models of Intercity Freight Movement.* Northwestern University.

Kohler-Reed, Susan, ed. 1994. *1995 Caribbean Basin Commercial Profile.* Cayman Islands: Caribbean Publishing Company Ltd.

Kullman, B. C. 1973. *A Model of Rail/Truck Competition in the Intercity Freight Market,* Vol. 15. Prepared for the U.S. Department of Transportation, Federal Railroad Administration.

Kresge, David T., and Paul O. Roberts. 1971. *Techniques of Transport Planning: Systems Analysis and Simulation Models*, ed. John Meyer, Vol. 2. Washington, D.C.: The Brookings Institution.

Murthy, A. S. Narshimha, and B. Ashtakala. 1987. "Modal Split Analysis Using Logit Models," *Journal of Transportation Engineering* 113, no. 5 (September): 502-517.

Nash, C. A., and A. E. Whiteing. 1987. Modal choice: A Total Distribution Cost Approach. In *Freight Transport Planning and Logistics Proceedings of and International Seminar on Freight Transport Planning and Logistics Held in Bressanon, Italy,* edited by Lucio Bianco and Agostino la Bella.

National Transportation Atlas Data Bases. 1995. Bureau of Transportation Statistics, U.S. Department of Transportation, Washington, D.C.

Ortúzar, J. de D., and L. G. Willumsen. 1994. *Modeling Transport*, 2nd ed. New York: John Wiley & Sons.

Rand McNally Railroad Atlas. 1988. New York: Rand McNally and Co.

Roberts, Paul O., Moshe Ben-Akiva, Marc N. Terziev, Yu-Sheng Chiang. 1977. *Development of a Policy Sensitive Model for Forecasting Freight Demand, Phase One Report.* Cambridge: Center for Transportation Studies (CTS).

Surti, Vasant H., and Ali Ebrahimi. 1972. "Modal Split of Freight Traffic." *Traffic Quarterly.*

Terziev, Marc N., Moshe Ben-Akiva, and Paul O. Roberts. 1975. "Freight Demand Modeling: A Policy Sensitive Approach," presented at the 47th National ORSA Meeting. Cambridge: CTS.

Texas-Mexico Multimodal Transportation. 1993. LBJ Policy Research Project Report Number 104. By Leigh B. Boske, Robert Harrison, Chandler Stolp, and Sidney Weintraub, LBJ Policy Research Project Directors. The University of Texas at Austin.

The Texas Seaports and Inland Waterway System. 1995. LBJ Policy Research Project Number 114. By Leigh Boske, director.

Transmode Consultants, Inc. 1994. *Truck-Rail, Rail-Truck Diversion Model: User's Manual.*

Transportation Research Board (TRB). 1977. *Freight Data Requirements for Statewide Transportation Systems Planning: Research Report*, NCHRP Report #177.

TRB. 1983. *Application of Statewide Freight Demand Forecasting Techniques*, NCHRP Report #260.

TRB. 1977. *Freight Data Requirements for Statewide Transportation Systems Planning: User's Manual,* NCHRP Report #178.

TxDOT International Relations Office. 1995. *Aplicación de Sistemas de Información Geográfica en la Planeación del Transporte en México.* Fascimile from Secretaria de Comunicaciones y Transportes, March 7, 1995.

U.S.-Mexico Trade and Transportation: Corridors, Logistics Practices and Multimodal Partnerships. 1995. LBJ Policy Research Project Report Number 113. By Leigh B. Boske and Robert Harrison, LBJ Policy Research Project Directors. The University of Texas at Austin.

U.S. Custom House Guide. 1995. Philadelphia: North American Publishing Company.

Wilson, Patricia A. 1992. *Exports and Local Development: Mexico's New Maquiladoras.* Austin: University of Texas Press.

Winsor, Anita. 1994. *The Complete Guide to Doing Business in Mexico.* Atlanta: U.S.-Mexico Trade Network, Inc.

Winston, Clifford. 1983. "The Demand for Freight Transportation: Models and Applications," *Transportation Research* 17A, no. 6: 419.

World Almanac and Book of Facts 1996. 1996. Published by World Almanac, distributed in the U.S. by St. Martin's Press.

Zlatoper, Thomas J., and Ziona Austrian. "Freight Transportation Demand: A Survey of recent Econometric Studies," *Transportation,* no. 16(1) 1989/90: 27-46.

APPENDIX A.

COMPARISON OF REGRESSION RESULTS FOR ORIGIN COMMODITY SPECIFIC MODELS

Appendix A. Comparison of Regression Results for Origin-Commodity Specific Models

Machinery (Sch_B = 84)
East North Central (Origin = EC: Chicago)

Parameter	*Model 1* *Estimate*
INTERCEPT	-4.29
	-9.44
DISTANCE	-0.002763
	-2.74
VALUE-BY-RAIL	-0.017721
	-2.42
R-square	0.513
Adj R-sq	0.4521
F-Value	8.427
F Prob	0.0032

Machinery (Sch_B = 84)
East South Central (Origin = ES: Birmingham)

Parameter	*Model 1* *Estimate*
INTERCEPT	-1.14
	-0.533
DISTANCE	-0.00187
	-1.507
POPULATION	0.1504
	1.49
R-square	0.1712
Adj R-sq	0.0607
F-Value	1.549
F Prob	0.2446

Machinery (Sch_B = 84)
Mid-Atlantic (Origin = MA: Philadelphia)

Parameter	*Model 1* *Estimate*	*Model 2* *Estimate*	*Model 3* *Estimate*
INTERCEPT	4.037	4.2183	3.10631
	2.97	**3.5**	**2.39**
DISTANCE	-0.003319	-0.002643	-0.004285
	-5	**-4.4**	**-5.79**
VALUE-BY-RAIL	-0.2133	-0.225613	-0.19792
	-7.82	**-8.9**	**-7.267**
FIRMS		0.09285	
		6.8	
EMPLOYMENT			1.008688
			5.471
SHIPPERS	0.0564		
	5.892		
R-square	0.8621	0.8467	0.8467
Adj R-sq	0.8303	0.8114	0.8114
F-Value	27.091	23.94	23.94
F Prob	0.0001	0.0001	0.0001

Machinery (Sch_B = 84)
Mountain (Origin = MT: Salt Lake City)

Parameter	*Model 1* *Estimate*	*Model 2* *Estimate*	*Model 3* *Estimate*
INTERCEPT	-3.0942	-2.969	-5.078425
	-2.82	**-2.7**	**-9.764**
DISTANCE	-0.001731	-0.001865	
	-2.25	**-2.363**	
VALUE-BY-RAIL	-0.05526	-0.057837	-0.282475
	-1.43	**-1.498**	**-2.42**
FIRMS	0.023064		
	1.485		
POPULATION		0.090757	
		1.543	
EMPLOYMENT			0.34314
			1.703
VALUE PER MILE (RL)			20.354105
			2.344
R-square	0.3654	0.3738	0.3729
Adj R-sq	0.1924	0.203	0.2019
F-Value	2.112	2.189	2.18
F Prob	0.1568	0.1469	0.1479

Machinery (Sch_B = 84)
New England (Origin = NW: Boston)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	15.14415	7.967	3.9919
	2.5	**1.5**	**1.21**
DISTANCE		-0.001414	-0.003081
		-1.433	**-3.123**
VALUE-BY-RAIL	-0.359641	-0.23422	-0.15905
	-3.437	**-2.483**	**-2.71**
SHIPPERS	0.138653		
	3.129		
FIRM-SHIP-WARE		0.054005	
		2.25	
EMPLOYMENT			1.18829
			2.436
R-square	0.5319	0.5642	0.5881
Adj R-sq	0.4468	0.4334	0.4645
F-Value	6.249	4.315	4.759
F Prob	0.0154	0.0339	0.026

Machinery (Sch_B = 84)
South Atlantic (Origin = SA: Atlanta)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	0.947097	1.40623
	0.841	**0.979**
DISTANCE	-0.00459	-0.003742
	-8.675	**-6.445**
FIRMS		0.060077
		3.668
VALUE-BY-RAIL	-0.026735	-0.029686
	-6.132	**-5.367**
EMPLOYMENT	0.679792	
	4.287	
R-square	0.882	0.9024
Adj R-sq	0.8607	0.8815
F-Value	36.001	43.147
F Prob	0.0001	0.0001

Machinery (Sch_B = 84)
West North Central (Origin = WC: Sioux City)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	-1.8305	-3.46995
	-1.019	**-7.43**
VALUE-BY-RAIL	-0.154101	-0.138702
	-3.571	**-3.482**
FIRMS	0.019878	
	0.946	
R-square	0.4631	0.4311
Adj R-sq	0.3915	0.3955
F-Value	6.469	12.125
F Prob	0.0094	0.0031

Machinery (Sch_B = 84)
West South Central (Origin = WS: Dallas)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	-3.026626	-3.0147	-2.713884
	-4.506	**-4.591**	**-3.987**
DISTANCE	-0.005203	-0.005573	-0.005347
	-4.857	**-5.265**	**-4.766**
SHIPPERS	0.019846		
	1.778		
POPULATION		0.123391	
		1.966	
R-square	0.5852	0.5974	0.5196
Adj R-sq	0.5437	0.5572	0.4968
F-Value	14.108	14.84	22.716
F Prob	0.0002	0.0001	0.0001

Machinery (Sch_B = 84)
West (Origin = WT: San Francisco)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	-3.237	-2.7462
	-2.966	**-2.666**
DISTANCE	-0.001669	-0.001966
	-2.476	**-3.039**
VALUE-BY-RAIL	-0.165494	-0.17034
	-4.252	**-4.624**
FIRMS	0.051248	
	3.3	
POPULATION		0.208702
		3.791
R-square	0.6497	0.6888
Adj R-sq	0.5879	0.6338
F-Value	10.51	12.54
F Prob	0.004	0.0001

Electronics (Sch_B = 85)
East North Central (Origin = EC: Chicago)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	-3.7515	1.884625	-0.035225
	-9.015	**1.078**	**-0.02**
DISTANCE	-0.003603	-0.002742	-0.003395
	-3.787	**-3**	**-3.095**
EMPLOYMENT			0.56579
			2.238
VALUE PER MILE (RL)	8.66675		
	3.361		
VALUE-BY-RAIL	-0.31375	-0.335562	-0.296754
	4.342	**-6.493**	**-5.293**
FIRMS		0.090661	
		3.376	
R-square	0.7599	0.7609	0.6805
Adj R-sq	0.7084	0.7096	0.612
F-Value	14.768	14.848	9.938
F Prob	0.0001	0.0001	0.0009

Electronics (Sch_B = 85)
East South Central (Origin = ES: Birmingham)

Parameter	*Model 1 Estimate*
INTERCEPT	-3.231444
	-6.219
DISTANCE	-0.002614
	-2.997
VALUE-BY-RAIL	-0.084175
	-2.694
R-square	0.5128
Adj R-sq	0.4378
F-Value	6.841
F Prob	0.0093

Electronics (Sch_B = 85)
Mid-Atlantic (Origin = MA: Philadelphia)

Parameter	*Model 1 Estimate*
INTERCEPT	-3.296354
	-7.793
DISTANCE	-0.00421
	-3.12
VALUE-BY-RAIL	-0.056445
	-3.308
R-square	0.6717
Adj R-sq	0.612
F-Value	11.253
F Prob	0.0022

Electronics (Sch_B = 85)
Mountain (Origin = MT: Salt Lake City)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*	*Model 4 Estimate*
INTERCEPT	-6.115933	-5.99681	-3.257487	-3.617539
	-10.082	**-10.576**	**-3.572**	**-3.607**
DISTANCE			-0.002778	-0.002187
			-3.229	**-2.451**
POPULATION	0.177324		0.205205	
	2.372		**3.136**	
SHIPPERS		0.031163		0.028634
		2.548		**2.463**
VALUE-BY-RAIL	-0.392575	-0.396969	-0.248954	-0.238258
	-3.298	**-3.46**	**-2.939**	**-2.485**
VALUE PER MILE (RL)	1.323647	1.303414		
	2.304	**2.374**		
R-square	0.5259	0.6106	0.7009	0.6209
Adj R-sq	0.4306	0.4646	0.5888	0.4787
F-Value	3.773	4.182	6.25	4.367
F Prob	0.0591	0.0469	0.0172	0.0424

Electronics (Sch_B = 85)
New England (Origin = NW: Boston)

Parameter	*Model 1 Estimate*
INTERCEPT	-4.102158
	-8.393
DISTANCE	-0.004298
	-3.107
R-square	0.4458
Adj R-sq	0.3996
F-Value	9.652
F Prob	0.0091

Electronics (Sch_B = 85)
South Atlantic (Origin = SA: Atlanta)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	-3.885537	0.0553	19.134427
	-9.652	**0.03**	**4.716**
DISTANCE	-0.004127	-0.003874	
	-4.096	**-4.419**	
FIRMS		0.063292	0.367813
		2.187	**5.655**
VALUE PER MILE (RL)			-0.50319
			-6.269
VALUE-BY-RAIL	-0.074404	-0.160771	-0.50319
	-1.837	**-3.05**	**-6.269**
R-square	0.659	0.7694	0.7973
Adj R-sq	0.597	0.7002	0.7365
F-Value	10.63	11.119	13.11
F Prob	0.0027	0.0016	0.0008

Electronics (Sch_B = 85)
West North Central (Origin = WC: Sioux City)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	-1.59615	-1.727012
	-1.825	**-1.856**
DISTANCE	-0.003039	-0.003023
	-2.672	**-2.474**
VALUE-BY-RAIL	-0.292068	
	-3.216	
VALUE PER MILE (RL)		-88.634667
		-2.753
R-square	0.5366	0.4713
Adj R-sq	0.4594	0.3831
F-Value	6.948	5.348
F Prob	0.0099	0.0219

Electronics (Sch_B = 85)
West South Central (Origin = WS: Dallas)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	-3.867853	-4.061815	-4.095
	-4.654	**-5.104**	**-5.18**
DISTANCE	-0.004	-0.004639	-0.004222
	-2.384	**-2.836**	**-2.673**
SHIPPERS			0.0208
			1.743
EMPLOYMENT		0.386626	
		1.65	
VALUE-BY-RAIL	-0.056657	-0.057982	-0.060432
	-1.746	**-1.886**	**-1.979**
R-square	0.3706	0.473	0.4828
Adj R-sq	0.2866	0.3601	0.372
F-Value	4.415	4.188	4.356
F Prob	0.0311	0.026	0.023

Electronics (Sch_B = 85)
West (Origin = WT: San Francisco)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	-3.6766	-3.636489
	-4.414	**-4.195**
DISTANCE	-0.002325	-0.002265
	-4.432	**-4.172**
POPULATION	0.182074	
	3.957	
SHIPPERS		0.030145
		3.664
VALUE-BY-RAIL	0.182074	-0.053897
	3.957	**-3.385**
R-square	0.7313	0.7093
Adj R-sq	0.6737	0.6471
F-Value	12.698	11.389
F Prob	0.0003	0.0005

Vehicles (Sch_B = 87)
East North Central (Origin = EC: Chicago)

Parameter	Model 1 Estimate	Model 2 Estimate
INTERCEPT	3.794784	2.088882
	1.839	**1.148**
DISTANCE	-0.00286	-0.003549
	-3.374	**-3.937**
EMPLOYMENT		1.122965
		4.34
FIRM-SHIP-WARE	0.058758	
	4.642	
VALUE-BY-RAIL	-0.060537	-0.043426
	-2.645	**-2.018**
R-square	0.7153	0.691
Adj R-sq	0.6496	0.6197
F-Value	10.886	9.691
F Prob	0.0007	0.0013

Vehicles (Sch_B = 87)
East South Central (Origin = ES: Birmingham)

Parameter	Model 1 Estimate	Model 2 Estimate
INTERCEPT	-4.680686	-4.375856
	-13.369	**-11.461**
EMPLOYMENT		0.623315
		3.673
FIRMS	0.212515	
	4.256	
VALUE PER MILE (RL)	-76.116108	-75.875155
	-2.904	**-2.591**
R-square	0.6717	0.6041
Adj R-sq	0.5987	0.5161
F-Value	9.206	6.866
F Prob	0.0067	0.0155

Veb cles (Sch_B = 87)
Mid Atlantic (Origin = MA: Philadelphia)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	1.58582	1.204157
	0.733	**0.51**
DISTANCE	-0.004375	
	-2.36	
POPULATION	0.246391	
	2.507	
FIRMS		0.208492
		1.891
VALUE-BY-RAIL	-0.415281	-0.393481
	-2.746	**-2.384**
R-square	0.5831	0.3883
Adj R-sq	0.4267	0.2524
F-Value	3.729	2.857
F Prob	0.0606	0.1095

Vehicles (Sch_B = 87)
Mountain (Origin = MT: Salt Lake City)

Parameter	*Model 1 Estimate*
INTERCEPT	-5.256571
	-7.223
DISTANCE	-0.00121
	-2.261
R-square	0.3173
Adj R-sq	0.2552
F-Value	5.112
F Prob	0.045

Vehicles (Sch_B = 87)
New England (Origin = NW: Boston)

Parameter	*Model 1* *Estimate*	*Model 2* *Estimate*
INTERCEPT	1.853217	1.853217
	0.909	**0.909**
DISTANCE	-0.005605	-0.005605
	-1.985	**-1.985**
EMPLOYMENT	1.113011	1.113011
	2.61	**2.61**
VALUE PER MILE (RL)	-154.66498	
	-2.716	
VALUE-BY-RAIL	1.414574	-0.387621
	4.467	**-2.716**
R-square	0.8643	0.8643
Adj R-sq	0.6609	0.6609
F-Value	4.248	4.248
F Prob	0.1964	0.1964

Vehicles (Sch_B = 87)
South Atlantic (Origin = SA: Atlanta)

Parameter	*Model 1* *Estimate*
INTERCEPT	0.095526
	0.046
DISTANCE	-0.003025
	-2.546
POPULATION	0.188737
	2.12
VALUE-BY-RAIL	-0.060276
	-1.958
R-square	0.5676
Adj R-sq	0.4497
F-Value	4.813
F Prob	0.0223

Vehicles (Sch_B = 87)
West North Central (Origin = WC: Sioux City)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*
INTERCEPT	6.427972	13.767331	12.801352
	2.15	**2.3**	**2.432**
DISTANCE	-0.001865	-0.002991	-0.00279
	-1.503	**-2.108**	**-2.074**
FIRMS	0.473324		
	3.431		
SHIPPERS		0.129687	
		2.931	
FIRM-SHIP-WARE			0.105713
			3.152
VALUE-BY-RAIL	-0.23695	-0.393758	-0.373223
	-3.323	**-2.963**	**-3.173**
R-square	0.6268	0.5594	0.5907
Adj R-sq	0.5025	0.4125	0.4542
F-Value	5.04	3.809	4.329
F Prob	0.0255	0.0517	0.0379

Vehicles (Sch_B = 87)
West South Central (Origin = WS: Dallas)

Parameter	*Model 1 Estimate*
INTERCEPT	-5.786712
	-9.338
DISTANCE	-0.004061
	-2.27
R-square	0.2326
Adj R-sq	0.1875
F-Value	5.153
F Prob	0.0365

Vehicles (Sch_B = 87)
West (Origin = WT: San Francisco)

Parameter	*Model 1* *Estimate*	*Model 2* *Estimate*
INTERCEPT	-4.120803	-4.228183
	-3.701	**-3.779**
DISTANCE	-0.002096	-0.001989
	-2.599	**-2.478**
EMPLOYMENT	0.596874	
	2.907	
FIRM-SHIP-WARE		0.025842
		2.833
R-square	0.4453	0.4346
Adj R-sq	0.366	0.3538
F-Value	5.619	5.38
F Prob	0.0162	0.0185

APPENDIX B.

COMPARISON OF REGRESSION RESULTS FOR COMMODITY-SPECIFIC, POOLED ORIGIN MODELS

Appendix B. Comparison of Regression Results for Commodity Specific, Pooled Origin Models

Machinery (Sch_B = 84)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*	*Model 3 Estimate*	*Model 4 Estimate*
INTERCEPT	-3.4066	-7.7187	-7.265602	-3.099
	-5.4	**-2.626**	**-3.044**	**-1.39**
DISTANCE				-0.002644
				-3.54
FIRMS				0.189871
				2.006
SHIPPERS		-0.033827		-0.10561
		-1.5		**-1.685**
POPULATION			-0.182464	
			-1.671	
VALUE-BY-RAIL	-0.05679	-0.06355	-0.060249	-0.035396
	-3.303	**-3.678**	**-3.63**	**-2.617**
VALUE PER MILE (RL)	-12.86823	-18.417429	-17.398804	
	-2.676	-3.093	-3.255	
R-square	0.3802	0.4458	0.4596	0.5478
Adj R-sq	0.3182	0.3583	0.3743	0.4473
F-Value	60133	5.094	5.387	5.451
F Prob	0.0084	0.0094	0.0075	0.0047

Electronics (Sch_B = 85)

Parameter	*Model 1 Estimate*	*Model 2 Estimate*
INTERCEPT	-3.152545	-3.078291
	-4.008	**-3.98**
DISTANCE	-0.002785	-0.002409
	-2.756	**-2.608**
FIRMS		
POPULATION	0.146706	
	1.892	
FIRM-SHIP-WARE		0.017619
		2.052
VALUE-BY-RAIL	-0.096646	-1.01438
	-2.832	**-3.003**
R-square	0.4667	0.4833
Adj R-sq	0.3667	0.3864
F-Value	4.667	4.989
F Prob	0.0158	0.0125

Vehicles (Sch_B = 87)

Parameter	*Model 1 Estimate*
INTERCEPT	5.145564
	1.899
DISTANCE	-0.003199
	-4.216
FIRMS	0.482133
	3.902
VALUE-BY-RAIL	-0.064839
	-2.405
R-square	0.6018
Adj R-sq	0.5271
F-Value	8.06
F Prob	0.0017