BUSINESS AND SITE SPECIFIC TRIP GENERATION METHODOLOGY FOR TRUCK TRIPS

Project 05-03
September 2005

Midwest Regional University Transportation Center
College of Engineering
Department of Civil and Environmental Engineering
University of Wisconsin, Madison

Authors: Kazuya Kawamura, Hyeon-Shic Shin, Sue McNeil
Urban Transportation Center, College of Urban Planning and Public Affairs, University of Illinois, Chicago
Libby Ogard
Prime Focus LLC

Principal Investigator: Kazuya Kawamura
Assistant Professor, Urban Planning and Policy Program, University of Illinois, Chicago

EXHIBIT B

Technical Report Documentation Page

<table>
<tr><td colspan="2">1. Report No.</td><td colspan="2">2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.

CFDA 20.701</td></tr>
<tr><td colspan="4" rowspan="2">4. Title and Subtitle
Business and Site Specific Trip Generation Methodology for Truck Trips</td><td colspan="2">5. Report Date September 30, 2005</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="4">7. Author/s Kazuya Kawamura, Hyeon-Shic Shin, Sue McNeil, and Libby Ogard</td><td colspan="2">8. Performing Organization Report No.
MRUTC 05-03</td></tr>
<tr><td colspan="4" rowspan="2">9. Performing Organization Name and Address
Midwest Regional University Transportation Center
University of Wisconsin-Madison
1415 Engineering Drive, Madison, WI 53706</td><td colspan="2">10. Work Unit No. (TRAIS)</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTRS 99-G-0005</td></tr>
<tr><td colspan="4" rowspan="2">12. Sponsoring Organization Name and Address

U.S. Department of Transportation
Research and Special Programs Administration
400 7th Street, SW
Washington, DC 20590-0001</td><td colspan="2">13. Type of Report and Period Covered
Final Report [06/01/04 – 09/30/05]</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="6">15. Supplementary Notes
Project completed for the Midwest Regional University Transportation Center with support from the Wisconsin Department of Transportation.</td></tr>
<tr><td colspan="6">16. Abstract

The motivation for this research comes from the recognition that recent developments in supply chain management (SCM) have altered the mechanism of truck trip generation at the individual facility level. This research develops models of truck trip generation (TTG) at the disaggregate level that incorporate strategic supply chain decisions made by individual businesses. The main assumption is that the TTG is an outcome of a series of strategic and operational business decisions. The research team conducted a survey of national retail chains. The data sets obtained from two furniture chains were used to develop binary logit models. Empirical data, although limited, validated the potential of building a disaggregate TTG model at the individual store level. Inclusion of location and store type dummy variables almost always improved model's predictive power, often dramatically. The findings presented in this report also underscore various shortcomings of existing methods. We found that commonly used independent variables such as the store floor space or the number of employees are poor predictor of truck trip generation at retail stores.</td></tr>
<tr><td colspan="2">17. Key Words</td><td colspan="4">18. Distribution Statement
No restrictions. This report is available through the Transportation Research Information Services of the National Transportation Library.</td></tr>
<tr><td colspan="2">19. Security Classification (of this report)
Unclassified</td><td colspan="2">20. Security Classification (of this page)
Unclassified</td><td>21. No. Of Pages
107</td><td>22. Price
-0-</td></tr>
</table>

Form DOT F 1700.7 (8-72) **Reproduction of form and completed page is authorized.**

DISCLAIMER

This research was funded by the Midwest Regional University Transportation Center through the Transportation Asset Management Pooled Fund Research Program administered by the Wisconsin Department of Transportation and the Federal Highway Administration. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the Midwest Regional University Transportation Center, the University of Wisconsin, the Wisconsin Department of Transportation or the Federal Highway Administration at the time of publication.

The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of the document.

ACKNOWLEDGEMENTS

We would like to thank following individuals who provided valuable input. We also would like to thank anonymous individuals who provided crucial data. Without their contribution, this research would not have been possible.

Larry Chaplin - Genco

Steve Silvis - Genco

Michelle Gripenstraw - Genco

Rober G. Bremley - Schneider Logistics

Kouros Mohamadian - University of Illinois, Chicago

Anthony Pagano - University of Illinois, Chicago

Paul Metaxatos - University of Illinois, Chicago

F. Gerald Rawling - Chicago Area Transportation Study

Keith Sherman - Illinois Department of Transportation

Bob Gale - Minnesota Department of Transportation

David Cipra - Wisconsin Department of Transportation

David Levinson - University of Minnesota

Table of Contents

EXECUTIVE SUMMARY

Introduction

The growth of E-commerce and increasingly sophisticated supply chain management strategies used by today's businesses require truck travel demand forecasting tools that are capable of capturing the effects of those market and economic forces on trucks' trip-making behaviors. As the first step toward the development of such model, this study tackled the most fundamental but often neglected component of truck travel demand forecasting process, trip generation. Our effort focused on building prototype models for one specific type of facility, retail stores.

Truck trip generation (TTG) analysis is a study to estimate the number of trucks coming in and out of a study area (e.g., a store, a shopping mall, or an industrial park). Thus, the TTG analysis provides transportation planners and public agencies with fundamental information, namely the usage of infrastructure in the vicinity of various businesses by trucks. This information is useful, for making transportation asset management decisions. Our approach for developing the new generation of TTG modeling is founded upon the observation that in order to capture the effects of supply chain strategies, it is necessary to construct a model at the individual facility level as opposed to at zonal level. In addition, it is necessary to identify the variables (preferably observable) that can be used to capture the characteristics of supply chain strategies employed at each facility.

Conceptual Model of Supply Chain and Truck Trip Generation

Figure E-1 is a conceptual supply chain of a manufacturer that was developed based on the information gathered from literature in the business and logistics fields. The supply chain consists of the manufacturer, suppliers, and consumers. Suppliers can be providers of raw materials, or maybe other manufacturers, value added assemblers, importers, or others. Consumers may include the end-consumers, Distribution Centers (DCs,) retailers, and others. The box in the center represents the logistics decision-making process. Strategic, tactical, and operational decision makings comprise the logistics system (Winser, 2003; Miller and de Matta, 2003). Based on the decision hierarchy of the logistics system, a business makes various decisions on inventory, distribution, production, sales, and replenishment or routing schedule to retail chains. However, businesses must be responsive to their customers and market demand.. Their decisions are partly dictated by the factors that they do not control. For example, businesses must respond to the conditions of various markets (e.g. fluctuations in consumer demand, seasonal demand spikes, labor, real estate, energy, and competitive pressures), consumer preferences, and government regulations. Consequently, truck trips connecting suppliers, firms, and customers are influenced by the complex interaction among those decisions, factors and constraints.

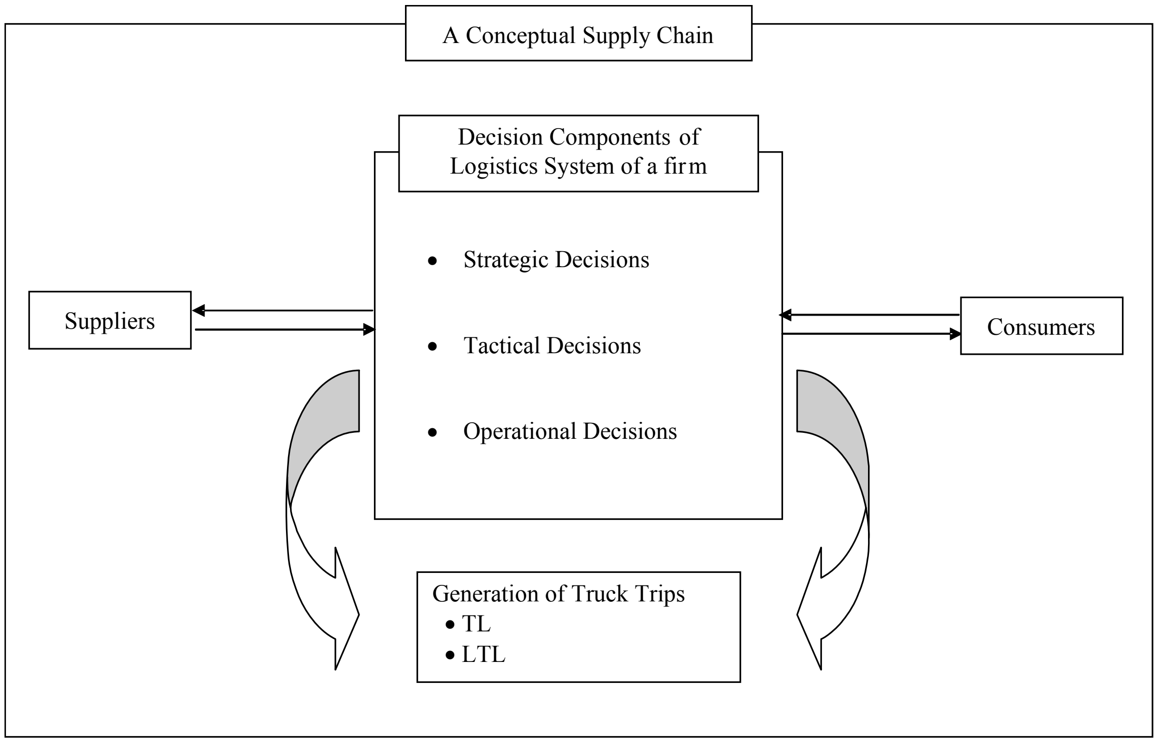


Figure E-1 Conceptual Supply Chain and Truck Trip Generation

Conceptual Truck Trip Generation Model

While the literature review provided a high-level understanding of the relationship between the business decisions, supply chain systems, and TTG, more detailed knowledge of the day-to-day operations of the businesses was needed to formulate actual model. In order to obtain needed information, preliminary interviews with the experts from a manufacturing plant, a trucking company, and two logistics and supply chain providers were conducted. Following is a list of the insights obtained from the interviews:

- All interviewees agreed that a disaggregate TTG analysis at the site-specific level would capture the activities of individual facilities.

- The experts recommended the retail sector, big-box stores in particular, as these facilities have sufficient volumes and activities where consistent trucking operations could be observed or estimated. Especially, the distribution centers (DCs) for those stores can be a rich data source for the activities of individual stores. The economic significance of the retail sector also makes these centers desirable subjects.
- Grocery stores were eliminated as the possible target for this research since, unlike big-box retailers, their supply chains are highly decentralized, making the data collection difficult if not impossible.
- In general, major retailers operate their shipping and routing schedules based on highly standardized supply chain management schemes so that different sectors may share certain characteristics.
- The number of employees and facility size are possibly important proxies for estimating the number of truck trips from a retail store.
- One of the interviewees stated that the most important variable is the customer demand, and logistics decisions are often in response to projected sales.

Assimilating the knowledge from literature review and expert interviews, the resulting framework for the retail businesses, depicted in Figure E-2 was developed. The small box on the top of the figure implies that the most important factor of freight truck demand is consumer demand for different types of goods. There are two relevant aspects of goods that can be used to categorize them in terms of TTG. The first is the velocity of inventory turns (e.g. slow-moving vs. fast-moving). The other is the weight-to-volume ratio. The goods that fill the capacity of a truck in terms of weight is called "weigh-out" goods, while those that face limitation from the cargo space is called "cube-out" goods. Today as a result of packaging characteristics most loads moving to retail stores cube out before they weigh out. It is typical that raw materials such as lumber, cement and bulk liquid chemicals weigh out before they cube out. The bottom of the middle box shows the variables that need to be considered in a TTG modeling process. The variables are classified as either long-term or short-term factors. The long-term factors include variables such as physical constraints of a facility and human resources. On the other hand, the short-term factors are associated with daily operations of a business. Such variables as replenishment schedule and sales information can be critical factors that determine the TTG.

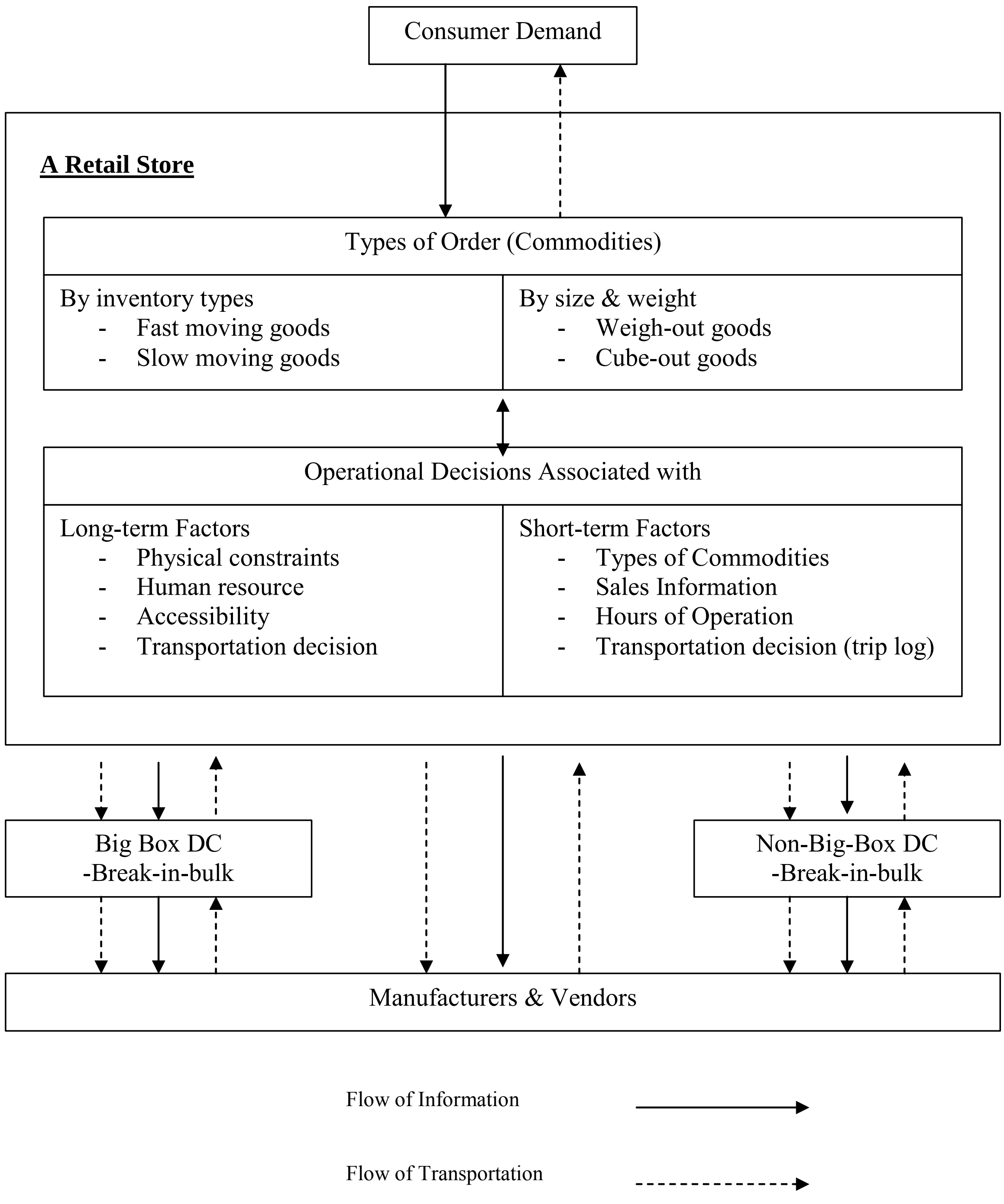


Figure E-2 Conceptual Model of Truck Trip Generation

Data Collection

Once the conceptual model was developed, the data for calibration and validation of the TTG model were collected. Due to the resource constraints and also following the recommendation made during the expert interviews, the data collection focused on the DCs and retail stores in the Midwestern states. Three survey methods were employed: a survey questionnaire, store visits, and phone contacts. The two data sets from the first provided the most detailed information on the amount of merchandise delivered to stores and the number of delivery trips per week per store. In total, 426 stores from 9 national retail chains – 5 furniture chains and 4 shoe chains – were collected, as shown in Table E-1.

In addition, for each store in the database, a total of 37 socioeconomic variables were compiled from the Census Block Groups that are within a 6-mile radius from the store. These variables were considered to be the proxies for the consumer demand for each store. Four different types of weights, based on the distance between each Census Block Group and the store, were applied to the socioeconomic variables.

Exploratory Data Analysis

Prior to the model development, a series of exploratory analyses, using descriptive statistics, was conducted. Several relevant observations are:

- All the retailers in the database have highly standardized routing schedules throughout the year. Except for furniture chains A and E, other seven chains have the same number of deliveries for all the stores throughout the year (see Table E-2).

- For furniture chains A and E, some stores receive only one delivery per week while others receive two deliveries per week. Such standardization can be made possible by consolidating shipments and sourcing all items from a single distribution center..

There is evidence that some furniture retailers are shifting from push to pull-logistics. For example, we found that Furniture Chain C recently changed from twice-a-week replenishment using tractor-trailer units to daily replenishments using a parcel carrier. All the large furniture pieces are delivered directly to the customers from the DC

Table E-1 Survey Results

Strategy	Chains	SIC*	Response Rate	Number of Stores	Advantage	Disadvantage
Distributing Survey Questionnaire to Distribution	Furniture A	5712	4% (3/75)	76	* Detailed replenishment	* Extremely low response rate * Time consuming * No consistent information
	Shoe A	5661		259		
	Apparel A	5632		n/a		
	Subtotal			**335**		
Visiting Stores	Furniture B	2511	100% (4/4)	4	* Cooperative	* No detail information * Time consuming
	Furniture C	5719	100% (6/6)	6		
	Furniture D	2512	100% (6/6)	12		
	Subtotal		100% (16/16)	**22**		
Phone Survey (Individual Stores)	Furniture E	5719	26.1% (13/36)	13	* Somewhat responsive * Low cost	* No detail information
	Shoe B	5661	67.5% (27/40)	27		
	Shoe C	5661	50% (10/20)	10		
	Shoe D	3149	79.2% (19/24)	19		
	Subtotal		59% (69/117)	**72**		
Total				**426**		

* The definitions of 4-digit SIC
5712 - Furniture Stores
5719 - Micellaneous Home Furnishings Stores
2511 - Wood Househod Furniture, except upholstered
2512 - Wood Household Furniture, upholster
5661 - Shoe Stores
3149 - Footwear, Except Rubber, Not Elsewhere Classified

Table E-2 Replenishment Schedules

Business	**Replenishment deliveries per week**	**Truck used**	**SIC**
Furniture Chain A	1 or 2	Semi	5712
Furniture Chain B	1	Semi	2511
Furniture Chain C	5	UPS	5719
Furniture Chain D	1	Semi	2512
Furniture Chain E	1 or 2	Semi	5719
Shoe Chain A	1	Semi	5611
Shoe Chain B	2	Semi	5611
Shoe Chain C	1	Semi	5611
Shoe Chain D	1	Semi	3149

- Standard industrial classification (SIC) is not a reliable scheme to categorize businesses in terms of replenishment frequencies. This implies that in order to develop a set of TTG models that covers entire spectrum of businesses that generate truck trips, a new scheme to categorize them in terms of TTG characteristics must be developed (so that one TTG model can be used for all the businesses in each category). In Table 2 businesses that were similar would categorized differently by SIC code.

According to the data source for Furniture Chain A, the most relevant determinants of the replenishment frequency was the type and location of each store. In general, the stores located in off-mall setting (i.e. customers can park directly in front of the stores) received more frequent replenishment than their mall-based counterparts due to their higher volume of sales and probably larger show room floor space. In addition, the "combo" stores that sell both regular and children's furniture tend to receive more frequent deliveries than the conventional stores. The analysis of replenishment frequencies for the Furniture Chain A stores generally supports such patterns although the store type (i.e. combo, conventional, or outlet) seems to have greater influence on the replenishment frequency than the location (i.e. mall or off-mall).

Model Building

The TTG model was developed using data from the two retail chains, Shoe Chain A and Furniture Chain A, that provided the most detailed information. The objective of the model building process was to answer the following questions:

- What are the variables that can be used to identify different supply chain strategies?
- What are the variables that can be used to predict customer demand at each store?
- What variables are the most appropriate independent variables for TTG model?
- Are TTG models fitted for one business, transferable to another business in the same retail sector?

The model fitting techniques used in this study were rather unusual due to the fact that the datasets that were available for model building, although they were very detailed, came only from two businesses. This is somewhat analogous to having a panel dataset that contains very detailed information for only one subject. It is not possible to estimate the inter-subject variability based on the standard statistical techniques such as standard errors and hypothesis tests. Thus, the model building relied entirely on the goodness of fit measures such as pseudo-R-squared, and percent of correctly predicted responses, as well as the reasonableness of the coefficients.

For Shoe Chain A, it was not possible to develop a TTG model since all the stores received 1 truck trip per week. Instead, a model that predicts the number of shoe cartons delivered per week to each store was developed to identify the variables that can be used to estimate the customer demand. Surprisingly, socioeconomic characteristics of the market shed were found to be poor determinants of customer demand. While the number of employees at each store was identified as the most important independent variable for estimating the customer demand, in general, the models generally showed poor levels of fit for the customer demand.

The TTG model for Furniture Chain A was fitted with binomial logit regression. The two best models, as shown in Table E-3, with different sets of independent variables and weights for socioeconomic variables, showed adequate level of fit and generally replicated the replenishment frequencies of the stores included in the fitted data[1]. However, both models failed to identify a large portion of the stores that receive two replenishment deliveries per week. This will lead to an underestimation of truck trips generated.

The two best models were applied to the data for Furniture Chain E for validation and to test the transferability. Since Furniture Chain E does not have the same type of store definitions as Furniture Chain A, a variant of the two models shown in Table E-3 was developed. The results showed that the best model, the variant of Model 22, correctly predicted the delivery frequency for 64% (7 out of 11) of stores. However, the model correctly estimated only 40% (2 out of 5) of stores that receive only one delivery per week. It should be noted however, that the fit of these variants were considerably poorer than the original models. For example, for Model 22, the pseudo R-squares dropped from 0.5076 to 0.2182 when the store type was removed from the model. This underscores the importance of incorporating the independent variables into the TTG model that are able to capture the supply chain strategies of businesses.

[1] The dependent variable for the TTG models was the frequency of deliveries/replenishments. Each delivery/replenishment generates two truck trip ends (attraction and production).

The findings from the model building generally supports the conceptual TTG model for big-box retail chains, shown in Figure E-2, except for the role played by the customer demand. Including the variables that capture customer demand, the number of employees or sales, in the models had little effect on the performance. In some cases, dropping those variables actually improved the model's ability to predict the stores with twice-a-week replenishment schedules. On the other hand, the role of the store type and location, key factors of supply chain decisions according to a survey respondent, turned out to be even more significant than our initial expectation. In fact, it is possible to correctly predict delivery frequencies for over 85% of the stores in the Furniture Chain A dataset using only the store type and location dummy variables.

Table E-3 Model Fit and Performance

	Model	
	Model 21	**Model 22**
N	58	58
Variables	employee, off-mall dummy, outlet dummy, conventional store dummy, pop. in high income grp.*, median age*, median income^2*	sales, off-mall dummy, outlet dummy, conventional store dummy, pop. density*, pop. in low income grp.*, median age*, total income^{-1}*
Pseudo-R^2	0.5602	0.5076
2 delivery stores (% correct)**	57.14%	64.29%
1 delivery stores (% correct)**	97.73%	100%
Overall % correct**	87.93%	91.38%

* weighted

** for calibration dataset (Furniture Chain A)

Implications

This research proposed a new generation of truck trip generation (TTG) modeling. Unique features of the proposed model, compared against existing TTG models, are: 1) it was developed at individual facility level, and 2) it was designed to capture the effects of supply chain strategies on the truck trip generation.

The knowledge gained during the data collection process brought valuable insights into the relationship between the business strategies concerning logistics and supply chain management and the TTG characteristics of businesses. The knowledge building process also taught us that a closer cooperation between the private sector and the academia is absolutely critical for the development of better analytical tools to address increasing truck trips in urban areas.

The findings presented in this report underscore various shortcomings of existing methods. As shown in Table E-4, the ITE method severely overestimated the truck trip ends while our TTG models produced reasonably accurate estimates.

Table E-4 Comparison of TTG Models Versus ITE Trip Generation Method

	Furniture Chain A	Furniture Chain E
Number of stores	58	11
	Total number of truck trip ends per week	
Model 23 (I-1)	126	34
Model 24 (II-1)	124	38
Model 21 W/O types (I-4)	132	40
Model 22 W/O types (II-4)	126	42
ITE (based on number of employees)	2265	588
Actual	144	34

We found that commonly used independent variables such as the store floor space or the number of employees are poor predictor of truck trip generation at retail stores. Although our study covered only two retail chains, this was consistent for both cases. Consequently, it is reasonable to suspect that current traffic studies and infrastructure planning activities that rely on such independent variables contain a large margin of error. Furthermore, for small-scale traffic studies, collecting such information may be rather wasteful since one can predict the TTG potential of a facility by simply asking the supply chain strategies of prospective tenant or collecting data on similar facilities operated by the tenant.

We also found that TTG characteristics can vary considerably within the same retail sector, e.g. furniture stores, depending on the supply chain strategy adopted. Since existing methods, e.g., ITE trip generation, categorize businesses based on the commodities or services being offered, this can introduce another source of error. Thus, for advancing the new-generation TTG model to the application stage, the development of appropriate classification system is imperative.

Recommendations

The findings from this study suggest a future direction for TTG modeling. Empirical data, although limited, validated the potential of building a disaggregate TTG model at the individual store level. Inclusion of location and store type dummy variables almost always improved model's predictive power, often dramatically. It should be noted that store location and types are physical characteristics that are easily observed, while identification of attributes such as floor space, the number of employees, or sales is more difficult to obtain. Since current data collection approaches that rely on land use information or development plans are not effective in gathering the aforementioned type of data, a whole new data collection strategy must be developed to support new TTG models.

The successful development of the new generation of TTG models will rest on the availability of data. Although a considerable amount of resources were spent on data collection, we did not obtain sufficient number of datasets to cover broad types of retailers. Also, the validation and the evaluation of the transferability of the model were limited by the data availability. Considering the proprietary nature of the data required to build the TTG models, it is our opinion that the development of the new generation models cannot be carried out without strong support from the public sector and also industry participants and trade organizations. We were fortunate to receive support from organizations such as the Council of Supply Chain Management Professionals to identify best practices in the industry. If individual companies could provide neutral trade associations or third parties with truck trip data a more robust analysis could be more meaningful.

1 . INTRODUCTION

1.1 Introduction

The fundamental motivation for this research is to understand the impacts of recent changes in supply chain strategies on the behavior of trucks. There are three broad trends that make freight transportation planning for trucks an urgent agenda in the U.S. First, economic trends in recent years seem to put ever-increasing number of trucks on roads. For example, the globalization of economy and the push toward free trade, exemplified by the North American Free Trade Association (NAFTA), have increased the role of freight transportation in border regions, major seaports, and corridors connecting main regions of production and consumption (Cambridge Systematics, 1997). Recent data show that trucking is the dominant mode of transportation for moving commodities in and out of major truck trip generators such as major urban areas, ports and border regions. For example, the proportion of urban interstates that carry more than 10,000 trucks per day on average is expected to increase to 69% by 2020 from 27% in 1998 (FHWA, 2004[a]). In addition, an estimate by the Federal Highway Administration (FHWA) predicts that the volume (tons) of freight transported by trucks will grow by over 75% in the next 15 years (FHWA, 2002). In the Chicago region, 75.3% or 220 million tons of commodities originating in the region ended within 50 miles of their origin; among those shipments, trucks shipped 90.3%, or 199 million tons (U.S. Bureau of Census, 1997). As the demands for faster and more flexible goods movements continue to grow, truck traffic will continue to increase. According to the 2002 Commodity Flow Survey (CFS), trucks move 64 percent of the nation's commercial freight, measured by value, and 58 percent of tonnage (BTS, 2004). Most certainly, such dependence on trucks has contributed in ever-increasing levels of congestion, road and bridge maintenance costs, and expansion needs for roads and freight facilities.

Second, in the early 1990's, freight transportation planning began to increased attention within the mainstream of planning activity, i.e. the development of regional transportation plans. This was partly influenced by the federal government's recognition of the importance of goods movement on the national economy. The Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 (US Congress, 1991) required that Metropolitan Planning Organizations (MPOs) include freight transportation components in MPOs' Regional Transportation Plans (RTPs), Transportation Improvements Plans (TIPs), and annual work elements. The Transportation Equity Act for the 21st Century (TEA-21) of 1998 (US Congress, 1998) followed its predecessor. Both pieces of legislation emphasized seamless goods movement (intermodality) and the efficient management of the National Highway System. It is expected that even greater emphasis will be put on freight transportation-related issues in the coming years.

Lastly, the widespread adoption of innovative logistics and supply chain management (SCM) strategies has changed the pattern of goods movements from push logistics to pull logistics. That is, the paradigm of goods movements has shifted from manufacturer- or supplier-led shipments (push logistics) of mass-produced items to consumer-led shipments (pull logistics). During the mass production era, most shipments were based on “point-to-point bulk shipments” (Suarez-Villa, 2003). In contrast, the current era of flexible production and lean inventory systems is characterized by more frequent goods movements with “smaller and lighter shipments” (Suarez-Villa, 2003). In the retail sector, for example, this leads to an increasing number of delivery trips without much change in the physical characteristics (e.g. floor space) of the stores.

1.2 Truck Trip Generation

While trucking plays a critical role in the national and regional economy, trucks are also responsible for most of the pavement damage, a sizable portion of air pollutants from non-stationary sources, and congestion (TRB, 2002). Thus, it is imperative that decision makers have accurate information on what drives the current and future demand for freight transportation in order to manage and utilize transportation infrastructure effectively.

Substantial strides have been made in forecasting truck travel demand in the past several years. However, the trends mentioned in the previous section will likely affect the pattern of freight flows especially in urban areas. It is evident that currently available demand forecasting methods are not suited to address those changes. A number of critical gaps need to be addressed.

As the first step toward the development of a truck demand forecasting model, that can account for the logistics and operations management strategies used by today's businesses, this study tackled the most fundamental but often neglected component of the truck travel demand forecasting process, trip generation. The truck trip generation (TTG) analysis is performed to estimate the number of trucks coming in and out of a study area or a facility. The analysis will help public agencies to make better transportation planning and policy decisions.

In practice; however, the TTG analysis has not often achieved its goal of providing accurate information for decision makers. The most serious obstacle is the lack of data. Accurate demand forecasting, and then the quantification of the impacts of truck activities can be computed only by using a reliable dataset (Holguin-Veras & Lopez-Genao, 2002). Data collection efforts in the public sector have been limited since the freight-related data often contain sensitive business information, and consequently, publicly available data are only released at the aggregate level (Pendyala, Shankar, & McCullough, 2000). Although disaggregate level data are available from private data collection companies like TRANSEARCH, the price of these datasets limits extensive use of this source in many regions. Because of these reasons, existing models for TTG (or truck travel demand in general) are mostly based on aggregate data. In other words, only aggregate variables or proxies of economic activities such as land use types, number of employees, and the gross floor space are used for TTG estimations. Consequently, existing models are not suited to analyze the impacts of logistics management strategies that may vary considerably among businesses.

In addition, the outputs produced by aggregate-level models are largely inconsistent; indicating such approaches cannot capture the relationship between economic activities at different business facilities and the amount of freight truck trips. This problem is exacerbated in today's environment as businesses have adopted even more sophisticated logistic management strategies. For example, flexible production decisions propelled by the development of just-in-time (JIT) logistics and e-commerce (or e-economy) must have changed the shipment patterns of goods.

This study presents a different perspective on the relationships between TTG and economic activities. Our hypothesis is that TTG is directly related to decision-making behavior with respect to supply chain management (SCM) and logistics strategies adopted by each business. At the site level, a retail store or a manufacturing plant for example, the number and type of freight truck trips within a given time period can be regarded as “an outcome of a series of decisions about products, sales, locations, delivery times, and frequencies (Iding et al., 2002).” As such, this study tries to relate the characteristics of the delivery trips made by trucks to both customer demand for merchandise and also strategic decisions that businesses make regarding their supply chain.

1.3 Research Questions

The lack data, due to confidentiality reasons prevented us from developing even rudimentary assumptions that can be used to formulate a set of candidate models to be tested. Thus, the research began with the exploratory analysis, followed by data collection and quantitative analysis. As such, the research questions, listed below, encompass a wide range of issues.

(1) What are the strengths and weaknesses of past TTG analysis?

(2) What are the businesses' perspectives on the mechanism that generates truck trips?

(3) How do TTG characteristics differ among businesses?

(4) How do the business operations relate to shipment decisions?

(5) What are the variables that can explain business operations?

(6) What are the most appropriate independent variables for predicting TTG?

(7) Are the TTG models transferable among similar businesses?

(8) Is there an existing taxonomy that can be used to group together the businesses for which a single model can be applied?

1.4 Organization of the Report

This chapter has discussed the background issues relevant to this research. In addition, research questions were formulated. The next chapter reviews various types of past TTG studies. The application of supply chain and behavioral aspects to freight transportation is also reviewed. Chapter 3 develops the research framework that is created based on the literature review and preliminary interviews of experts. Then, the data collection processes, that employed three different approaches, are discussed in detail in Chapter 4. After the description of the collected data in Chapter 5, exploratory analyses of the data are conducted in Chapter 6. Then, the TTG models are developed in Chapter 7. Finally, the report concludes with the discussion of the implications of the findings.

2 . LITERATURE REVIEW

The literature on estimating TTG can be divided into several categories depending on the unit of analysis, types of data, and objectives of research. An analysis of trip generation can be measured at either the regional level or the site-specific level. In addition, TTG analysis can be conducted either to evaluate or to predict travel behavior for a particular facility; or, TTG can be a part of a comprehensive transportation modeling system.

This chapter attempts to address the first research question stated in Section 1.3. Three aspects of TTG analyses are reviewed. First, the comparison between commodity-based and trip-based approaches is briefly summarized. Second, different types of TTG analysis are described in terms of various estimation techniques employed. Finally, the application of supply chain characteristics or behavioral aspects of freight routing decision is discussed.

2.1 Commodity-Based vs. Trip-Based Approaches

Freight transportation demand models can be loosely divided into two approaches depending on the study focus: commodity-based and trip-based approaches. The former focuses on the quantity and the types of goods movements by various types of vehicles, while the latter specifically deals with traffic flows in terms of the types of vehicles and operations (Garrido, 2001). However, the more practical distinction between the two approaches is the objective and the types of data used. The trip-based approach is usually applied at the site-specific level, while the commodity-based approach is applied at the zonal level. Since the zonal level economic data are easier to obtain than the site level information, the commodity-based approach is the one that is frequently used if the goal of the study is to forecast area-wide truck trips. One of the recent examples is the freight analysis framework (FAF) developed by the FHWA. The FAF is “a methodology to estimate trade flows on the Nation’s infrastructure, seeking to understand the geographic relationships between local flows and the Nation’s overall transportation system” (FHWA, 2004[b]).

The commodity-based approach does not directly estimate truck trips. The number of truck trips is calculated by “converting annual commodity tonnage into daily truck trips using a payload conversion factor” (Fisher & Han, 2001). Although the approach has been extensively studied and seems more popular than the trip-based approach, it tends to underestimate the number of truck trips generated within a region because the effects of trip chaining and local pickup and delivery trips are not accounted for (Fisher & Han, 2001). It also should be noted that most of the commodity-based models rely on simple conversion factors to estimate the truck trips from the commodity flow. Thus they are not capable of capturing the effects of logistics strategies on the load factors of trucks. In addition, the commodity-based approach requires commodity tonnage, which is not easy to obtain, except at freight terminals and ports.

The trip-based approach directly measures truck trips on the basis of the assumption that the number of truck trips produced in or attracted to an area is a function of some observable characteristics, such as the types of land use, number of employees, and gross floor space. Although the trip-based approach estimates the number of truck trips better than the commodity-based approach (Fisher and Han, 2001), it exhibits several problems: no methodological agreement, a lack of detailed truck flow data, the use of proxy variables at the aggregate level, and difficulty generalizing results.

2.2 Techniques of Truck Trip Generation Analysis

The following sections summarize the techniques for estimating TTG models. Some studies fall into two or more categories. An effort has been made to categorize these studies based on their stated focus.

2.2.1 Trip Rates

The calculation of trip rates is probably the simplest and the most straightforward technique of trip generation (Fisher & Han, 2001). This is generally a good approach for a short and mid-term impact analysis for a small study area (Brogan, 1979). In general, trip rates are expressed as the number of truck trips per land area or facility size in acres or square feet. Brogan (1979) measured land-area trip rates for Flint, Michigan; Columbus, Ohio; Kenosha, Wisconsin; and Racine, Wisconsin by various land-use classes. Although there was a large variation among the cities, it was generally recognized that commercial and industrial land uses are the most intense generators of truck trips in terms of trip rates per unit area. This study also compared the trip rates from those four cities to those from four previous studies. The study found that “the stability of the trip rates across urban areas is difficult" because a wide variety of land-use categories are employed. This may be due to the fact that the trip rates technique relies on a single independent variable (e.g. area).

2.2.2 Regression Analysis

Regression analysis seems to be the most frequently used method in trip generation analysis for passenger as well as freight transportation demand models. A regression-based trip-end model was developed by Slavin (1974) to identify the relationships between truck trip-ends and socio-economic activities at the various land uses. The trip-end model is characterized by the use of the information regarding trip purposes (activities) at the origin and destination zones. In Slavin's model, trip-ends were divided into several categories, based on different activities of industries, such as “production and distribution activities” and “delivery without production activities”. The study estimated the number of truck trips between origin and destination zones as a function of: the area of the destination zone, the number of employees for each industrial category, residential population, and travel times between origin and destination pairs. The study found that zonal trip generations, categorized by the trip-ends (activities), had statistically significant relationships with all independent variables. An important finding of this study is that “zonal trip making is a strong function of activity densities and proportional to zonal area.” (Slavin, 1974)

Brogan (1980) extended Slavin's work by comparing the effectiveness of different stratification strategies for estimating TTG by trip ends. This study employed three different stratification schemes for constructing the models: (1) by truck types, (2) by trip purposes, and (3) by land uses. Various independent variables were used including: highway employment, retail employment, manufacturing employment, dwelling units and others. Despite the efforts, the paper found that the stratification did not yield significant improvement over the non-stratified model.

A similar study was conducted in Fontana, California (Tadi & Balbach, 1994). The purpose was to measure trip generation characteristics for non-residential land uses, such as warehousing, industrial, truck terminals and truck sales. Data collection consisted of conducting 10-hour manual counts and 24-hour machine counts at 21 sites. The data were collected for two critical locations by 24-hour manual counts in order to ensure the statistical integrity of the data. Three regression models, two separate models by axle types and the pooled model, for each land use were constructed. Each model was a function of the building area or total site area. The statistical results were not satisfactory due to the small sample size.

Recent studies tend to focus more on the analysis at the site-specific level. Al-Deek et al. (2000) developed a trip generation model for predicting the levels of both inbound and outbound cargo truck traffic at the Port of Miami. The daily totals of truck trips were collected for a period of several months during 1996 and 1997. The model was estimated as a function of the number of imported or exported containers. The analysis result was statistically significant. In addition, a single equation autoregressive integrated moving average (ARIMA) time series model was developed. It was found that the model had the potential for success as a long-term forecast of the independent variables.

Holguín-Veras and López-Genao (2002) calibrated the TTG at marine container terminals based on a nationwide survey of 21 container terminals: eight in the North Pacific region, six in the North Atlantic region, five in the South Pacific region, and two in the South Atlantic / Gulf Coast region. Simple trip generation rates were computed as a function of the areas of container terminals. In addition, regression analysis was used to discover the relationships between the number of one-way truck trips and a set of independent variables such as the container volumes measured in both truck equivalent units (TEUs) and the number of containers. Regression models were developed for two different scenarios: typical days and busy days. The findings from the research were: (1) the number of boxes seemed to be the variable with the highest explanatory power, (2) the land use variables such as terminalarea and the number of berths were found to have only a modest to insignificant impacts. The analyses of both trip generation rates and the regression models indicated that there were statistically significant differences in the patterns of trip generation among the regions.

Fite et al. (2002) conducted a stepwise multiple linear regression analysis to forecast the freight volume for a truckload (TL) trucking firm – J.B. Hunt Transport, one of the world's largest carriers. The study used 107 broad economic and industrial indices, such as the Dow Jones utilities index, S&P 500 stock index, Consumer Price Index, and so on, as independent variables. Nearly three years of historical freight data from J.B. Hunt were collected and used as the dependent variable. The monthly freight data were regressed as a function of economic and industrial indices. The analysis was carried out for three levels: national, regional, and industrial. Although the use of economic indicators at the national or regional level did not provide statistically satisfactory results, the study recognized the need for a further research on the modeling of freight volumes based on broad economic and industrial factors.

2.2.3 Artificial-Neural Network

Al-Deek (2001) compared the model that was developed using the regression analysis in his previous research, (Al-Deek et al. 2000) to an artificial neural network (ANN) model with the application of back propagation neural networks (BPN). BPN is an artificial neural network (ANN) model that imitates the functions of human neurons. In BPN, the network consists of neurons (or nodes) and neuron synapses, or connections. It is simulated until the error function of input vectors is less than a preset tolerance.

The objective was to find an accurate model for trip generation and modal selection. The prediction of the number of inbound and outbound truck trips was measured using regression and BPN. The total daily inbound and total daily outbound freight truck trips were used as the dependent variables. The daily volumes of exported and imported freight containers were the independent variables. The model was able to capture the time lag between the inbound trips and outbound trips. For example, today's volume of outbound trips was related to the number of inbound trips three days ago. Various weights, reflecting the days of the week, were applied to the inputs to the BPN. According to t-statistics with the 95% confidence interval for the regression analysis and the BPN, the latter was much more accurate than the linear regression analysis.

In addition to the comparison of the two methods, the model created by the application of ANN was applied to three ports in Florida: the Port of Jacksonville, the Port of Tampa and Port Canaveral (Al-Deek 2001; Klodzinski & Al-Deek, 2003). The purpose was to test the transferability of the ANN modeling technique. At a confidence level of 95% with the data collected from field locations around the ports, the models were successfully developed and validated. As for the selection of appropriate independent variables, which is a critical issue in TTG, the BPN model automatically selected them. In addition, BPN model could detect complex relationships between dependent variables and independent variables. However, the major disadvantage of ANN is the lack of well-defined guiding rules for developing the network, its dependence on intuition in deciding the model's stopping criteria of simulation, the computational burden, and the requirements for the detailed data.

2.2.4 Origin-Destination Trip Matrix Estimation

Although trip matrix estimation has been an area of research for some time, the application to freight travel demand model is relatively new. If accurate data for trip matrices are available, it is possible to calculate the accurate trip generation for each point or zone of concern. However, the availability of data poses a significant challenge to researchers, requiring various estimation techniques on the basis of a combination of different data sources and freight attributes.

List and Turnquist (1994) developed a method for estimating multi-class truck trip matrices from three different data sources: (1) link volumes or classification counts, (2) partial OD estimates, and (3) cordon counts. Trip matrices were estimated for three truck classes: van, medium and heavy trucks. They concluded that the truck flow changes were related to the commodities being carried and the physical characteristics of trucks.

2.3 Supply Chain models

The purpose of reviewing supply chain management (SCM) strategies is related to the scope of this study. The study hypothesizes that TTG is an outcome of a series of strategic business decisions regarding thesupply chain. Thus, the knowledge of supply chain strategies will help the public sector understand the behaviors of businesses and provide insights into plan and policy developments.

Recent research efforts seem to indicate that the estimation of the number of truck trips has been expanded from the analysis of easily observable data to the incorporation of the behavioral components of business activities. This trend reflects the fact that the relationships between economic activities and transportation flows is not static; rather, these decisions are affected by various strategies of different industries (Iding et al., 2002). In other words, decision characteristics of businesses are latent variables that influence the truck trip generation. This section describes the concepts and applications of supply chain, logistics, and activity-based models to freight transportation.

Although detailed components vary by businesses and studies, Mentzer et al. (2001) likened a supply chain to a “pipeline” through which information regarding products, services, financial resources, demands and forecasts flow. Such flows play a role in the coordination of inter-firm decisions such as marketing, sales, research and development, forecasting, production, purchasing, logistics, information systems, finance, and customer service. Transportation management in a supply chain involves the choices of shipping schedules, modes, and timetables, with the objective to minimize the shipping costs (Gaither & Frazier, 1999). Various operations research methods, such as linear and non-linear programming, are used to optimize transportation decisions. The purpose is to minimize the overall cost of transportation, subject to the supply constraints of industries and the demand constraints of consumers (Winston, 1994).

Boerkamps et al. (2000) developed a conceptual model, “GoodTrip model,” to estimate freight movements. The model reflects the interactions among markets, actors, and supply chain elements of urban freight movement. The model begins by recognizing the lack of behavioral aspects of traditional freight demand models. That is, the authors try to incorporate the behavioral interactions between consumers’ demand on various goods and the responses of shippers, producers, carriers, and other freight-related interests in a supply chain. It was anticipated that considering these interactions would yield a more reliable model of a goods distribution system. The GoodTrip model is the application of the traditional four-step model to the supply chain at the scale that is between zone-based and disaggregated logistics models. The model begins with the estimation of consumer demand, followed by the estimation of goods flow, and the simulation of vehicle tours. Although the model has not yet been mathematically constructed yet, it provides an insight into the future development of freight demand models in that the changes in the supply chain will explain the behavior of freight flows.

Wisner (2003) modeled a supply chain using the structural equation technique that is a confirmatory approach to data analysis requiring a priori assignment of inter-variable relationships. A total of 5470 questionnaires were distributed to 1500 supply and material managers, 3000 senior managers, and 970 senior European managers of 1500 U.S. manufacturing firm. Based on 556 usable responses from the survey, four structural equation models were created: models for supplier management strategy, customer relationship strategy, supply chain management strategy, and firm performance. The assumption of the study was that the strategies and the variables related to the supplier, customer, and supply chain, influence a firm's performance. The model revealed the impacts of the variables of supply chain components on firm performance. However, as the authors admitted, it was not clear whether the model is statistically significant since "there is no single test of significance for structural equation models that can absolutely identify a correct model given the sample data" (Wisner, 2003).

Another interesting study is the application of the network equilibrium model to a supply chain. Nagurney et al. (2002) created a hypothetical network model of a supply chain consisting of two manufacturers, two retailers and two consumers. It was assumed that there is a flow of only one homogeneous product. The objective function was created to minimize the cost functions of firms and the handling costs of retailers. This model oversimplified the real world, yet provided insights with regard to the truck travel demand model. The cost function is the appropriate approach to estimate the flow of shipments between firms and retailers yet is usually ignored in TTG modeling.

The supply chain can be defined as the extension of logistics strategies of firms in which the logistics system consists of three decision hierarchies: strategic, tactical and operational (Winser, 2003; Miller & de Matta, 2003). The first two strategies involve the long-term (2 or more years) and mid-term (1-2 years) strategies of a firm. The operational strategy includes inventory, distribution, production, and transportation decisions. Since the operational strategy seems to provide detailed information of firm's activities, an in-depth review of this strategy should be carried out.

Further reviews of the research on the operational strategy suggested that there may be an explicit relationship between product and transportation. Most studies are based on network optimization, which tries to minimize transaction costs that include production, sales, and bargaining costs. The purpose of the network optimization is to identify the relationships between the capacity of a firm or warehousing, and management strategies (e.g. production costs and production rates) and transportation demand (Miller & de Matta, 2003; Khouha, 2003; Disney et al., 2003; Miranda & Garrido, 2004).

For Example, Miller and de Matta (2003) provided the conceptual relationships for the hierarchical decision structure of a firm; (1) at the strategic level, the firm creates strategies for its "profitability, growth, and market position," (2) this strategy is then reflected in determining the allocation of supply chain components such as distribution centers (DCs), production line capacity, and the location of the firm, and (3) finally, production and transportation schedules are decided. In addition, the multi-level manufacturing network consisting of two geographically separated plants was formulated and tested by integrating production and transportation schedules. The model was based on the evaluation of "major production, transportation, and inventory costs."

In the SCM studies mentioned above, transportation was treated as a cost component to be minimized while facilitating the flows of goods and services. However, decision variables of each firm could not be identified since the models treated each member of a supply chain as nodes within the whole network. Thus, dealing with the whole supply chain network is not desirable for identifying the decision variables for an individual firm that is a part of the supply chain.

2.4 Chapter Summary and Implications for the Study

In this chapter, several types of TTG techniques were reviewed.

(1) The difference between the commodity-based and trip-based approaches is the unit of measurement. Since the commodity-based approach does not directly measure the number of truck trips, the trip-based approach should be employed for this study.

(2) The trip rate method is a simple approach to measure trip generation. However, it is not reliable for the forecasting purpose since trip rates based on land area vary significantly from region to region and even within a region (Tadi & Balbach, 1994; Fischer & Han, 2001).

(3) Regression analysis was identified as the dominant approach for estimating TTG. Despite its popularity, there seems to be no agreement on the independent variables that should be used to estimate TTG. However, recent research at the site-specific level suggests that the analysis for a specific site can produce consistent results and can be transferred to other sites with similar characteristics. This suggests that a disaggregated level of analysis employing regression model may yield a good result.

(4) The application of the ANN produced better results than regression analysis (Al-Deek, 2001; Klodzinski & Al-Deek, 2003). However, the long process of trial and error, absence of defined rules for network design, and, consequently, the reliance on the researchers' intuition are problematic. In addition, the data requirement and measurement complexity are other obstacles. Nevertheless, if detailed data can be collected from individual businesses, ANN presents a potential.

(5) The estimation of O-D matrices tries to overcome the lack of data. Even though the disaggregate approach may exponentially increase the size of the matrices and, consequently, costs, time, and complexity of analysis, with accurate data, this approach may be appropriate for estimating trip generation for each point or zone of interest. However, as common in freight transportation planning, the availability of the data poses a significant challenge.

(6) In SCM, transportation is treated as a cost component to be minimized. Since a firm is considered as a node of a supply chain network, individual behaviors of a firm cannot be analyzed. In this sense, dealing with a whole network is not desirable for this study. On the other hand, incorporating behavioral aspects of logistics systems seems to provide an alternative for estimating truck trips. Although a substantial amount of detailed data are required, a more accurate analysis is probably possible. Despite the difficulty of collecting data and constructing a model, the promise of this approach lies in the fact that understanding business behavior will help the identification of appropriate set of independent variables for explaining the TTG.

3 . DEVELOPMENT OF RESEARCH FRAMEWORK

This chapter discusses the development of the conceptual framework of the TTG model. The discussion begins with the examination of the relationship between the characteristics of a supply chain and TTG in section 3.1. Section 3.2 provides the justification for conducting the TTG analysis at the disaggregate level. In section 3.3, the TTG analysis framework and assumptions used in this study are developed based on the information obtained through the literature review and preliminary interviews with field experts. This chapter concludes with a summary of the discussions in section 3.4.

3.1 Supply Chain and Freight Transportation

This study frames TTG based on the relationships between truck trips generated and economic activities. At the individual business level, the number and type of freight truck trips within a given time period can be regarded as an outcome of a series of business decisions about products, sales, locations, supply chain management, (SMC) (Iding et al., 2002) as well as social and market conditions in which the business operates. As reviewed in the previous chapter, past TTG studies mainly captured the effects of the latter factors using variables such as store sizes and number of employees.

A supply chain is a process that makes effective use of flows of information, such as material, production, sales and other decision-making processes, “starting with raw materials and ending with finished products delivered to the ultimate customer" (Gaither & Frazier, 1999). Figure 3-1 is a conceptual supply chain of a manufacturer, which consists of a manufacturer, suppliers, and consumers. Suppliers can be raw material providers, manufacturers, importers, and others. Consumers may include the end-consumers, DCs, retailers, wholesalers, and others. The box on the center is the logistics decision-making components of a business. Strategic, tactical, and operational decision-makings comprise the logistics system (Winser, 2003; Miller and de Matta, 2003). Based on the decision hierarchy of the logistics system, the business makes various decisions on inventory, distribution, production, sales, and replenishment or routing schedules to retail chains. The transportation component is a part of the firm’s profit maximization behavior in the supply chain. However, businesses do not operate in a vacuum. Their decisions are partly dictated by external factors. For example, businesses must respond to the conditions of various factor markets (e.g. labor, real estate, energy, and products and services being sold), consumer preferences, and government regulations. Consequently, truck trips connecting suppliers, firms, and customers are influenced by the complex interaction among those decisions, factors and constraints.

The effective management of the transportation component is a critical factor of supply chain performance (Bowersox et al, 2002). Since the planning and control of the transportation part of the supply chain critically affect the number of truck trips, estimating TTG based on activity-related variables of a firm's supply chain and logistics strategy will provide a comprehensive framework that is based on reality.

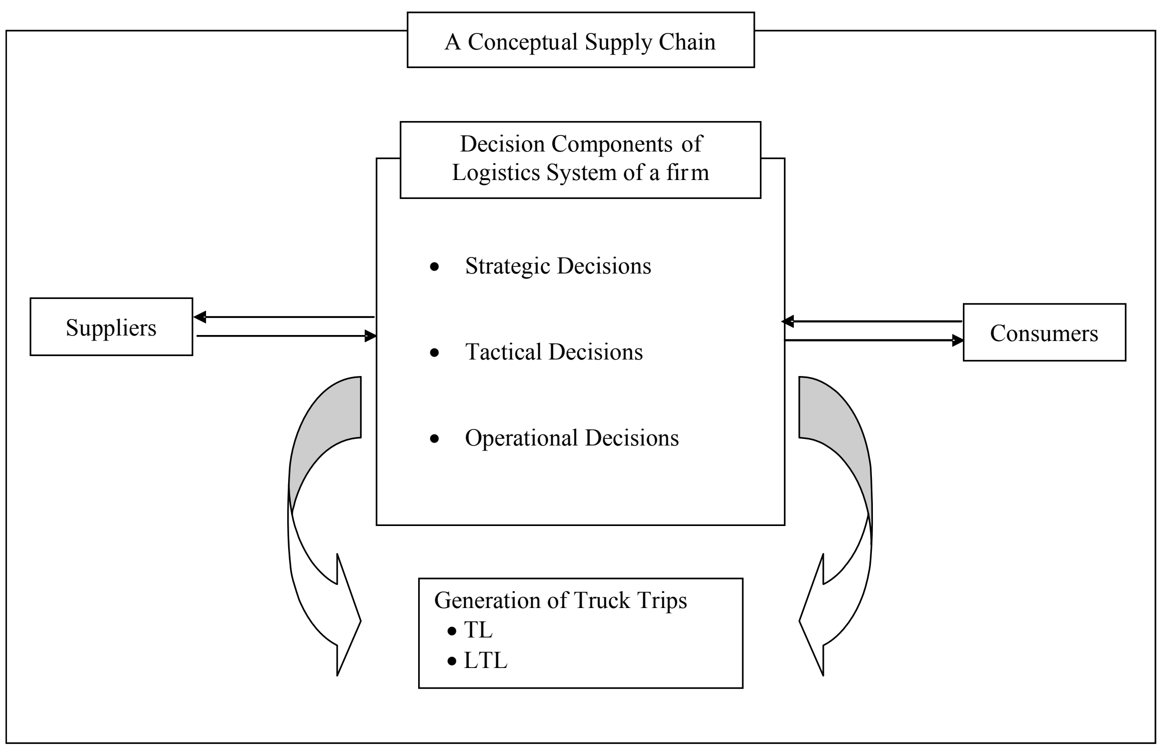


Figure 3-1 Conceptual Supply Chain and Truck Trip Generation

3.2 Justifications for the Disaggregate Approach

In the previous chapters, the lack of publicly available data and appropriate approach were identified as the most problematic issues in the past TTG studies. The use of spatially aggregated variables is prone to produce large aggregation errors since the models cannot capture heterogeneous characteristics of individual trip generators. In the manner that is analogous to the construction of the market demand curve in microeconomics, the demand model for truck trips should be constructed by aggregating the trips estimated at individual business level.

Since it is not practical to develop a TTG model for each business, even the disaggregated approach requires some type of aggregation scheme that identify the subjects with similar characteristics such that a single model can be applied for all the businesses within the same group. In other words, the models will be transferable among the businesses in the same group. Consequently, if it is possible to identify the proper taxonomy for forming the sectors that only include the businesses that share similar SCM strategies, the models will be transferable within each sector. It should be noted that the proper definition of the sector may not be related to the various industrial classification systems (e.g. Standard Industrial Classification (SIC), North American Industry Classification System (NAICS), etc.). As an example, Figure 3-2 and Figure 3-3 display two different supply chain networks for a retail chain that may share the same SIC or NAICS. Both figures describe the relationships between a retail store and its suppliers or DCs. In Figure 3-2, there are four types of inbound truck shipments associated with a retail store; two shipments are made from suppliers' DCs, one from a retailer-owned DC and a direct shipment from a supplier. On the other hand, a retail store in Figure 3-3 receives a shipment only from its own DC where all shipments by either less-than-truck-load (LTL) or by truck-load (TL) are consolidated and distributed. While both retail stores sell similar items which should lead to the same SIC classification, because of the way the parent company reports the business activities similar stores are actually identified differently. The relationship between stores and their suppliers cannot be reliably captured by the SIC or NAICS.

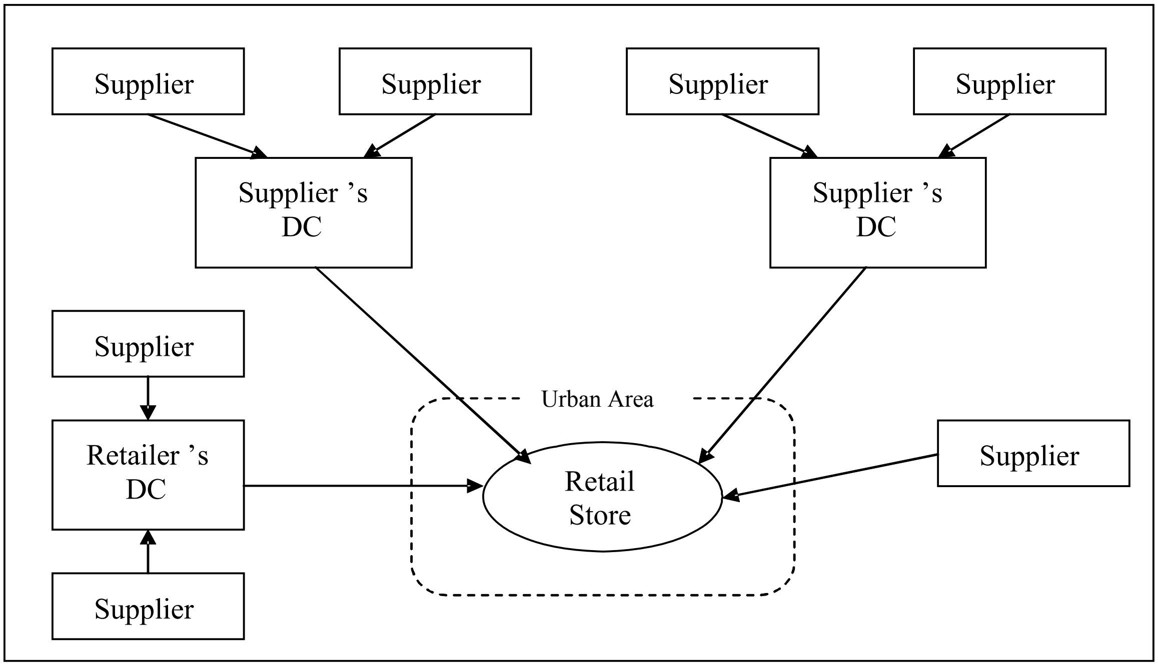


Figure 3-2 Supply Chain System and Truck Trip to a Retail Store- Example 1

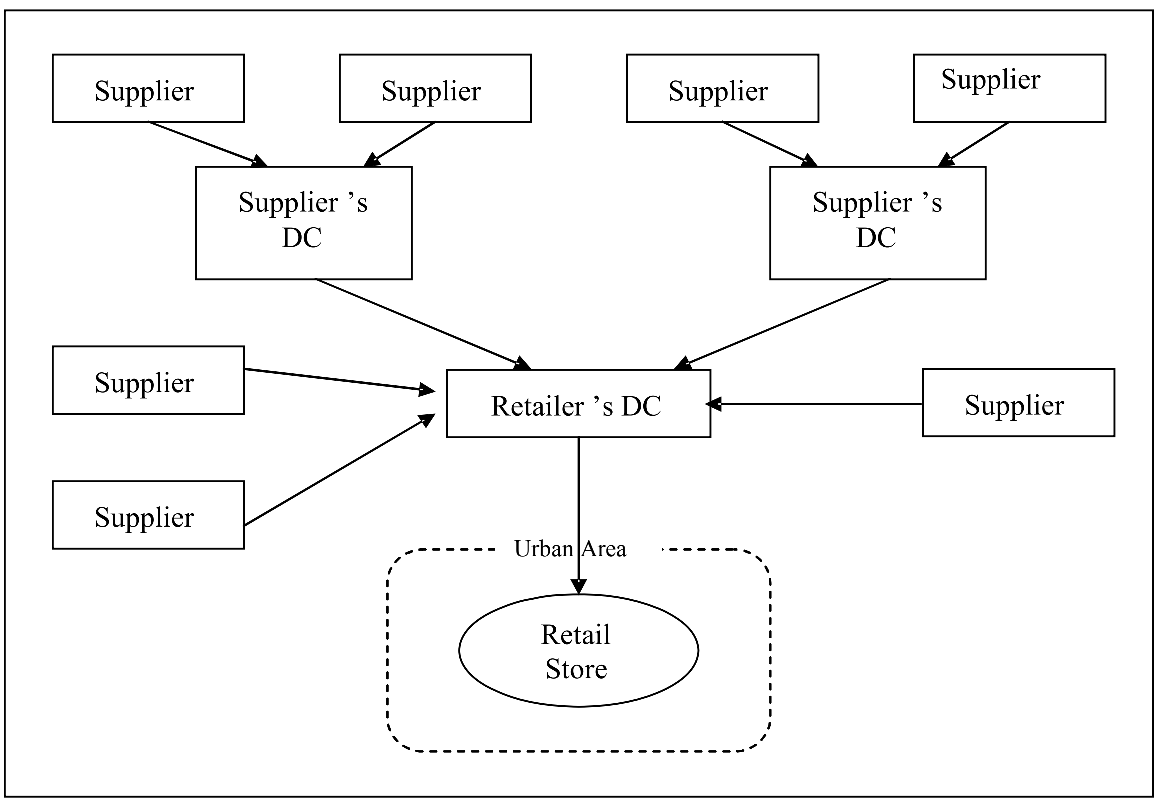


Figure 3-3 Supply Chain System and Truck Trips to a Retail Store - Example 2

3.3 A Conceptual Framework of TTG Model

While the literature review provided a high-level understanding of the relationship between the business decisions, supply chain systems, and TTG, more detailed knowledge of the day-to-day operations of the businesses were needed to formulate an actual model. In order to obtain needed information, preliminary interviews with the experts from a manufacturing plant, a trucking company, and two logistics and supply chain solution providers were conducted. The purpose of the interviews was to obtain feedback on our understanding of the relationship between the TTG and supply chain strategies, develop a model of business' decision-making process and the input factors related to TTG, identify possible data sources, and most importantly, determine industry sectors of interest for this research.

The experts recommended Distribution Centers (DCs) for big-box retail stores as the facilities at which would present the most consistent relationship between business operation and TTG. It was also noted that there would be sufficient volumes to observe. DC's can be rich data sources not only for their own trip generation, but also for the activities of individual big box retail stores because the DC controls and manages the delivery of merchandise to the final retail location. The retail sector provides daily necessities for consumers. Researching this sector provides examples that are easy for the public and planners to relate their own experiences to. It is also a large enough sector to have significant share of the American economy for study purposes.

The interviews provided other vital information. All interviewees agreed that a disaggregate TTG analysis at the site-specific level would be able to capture the activities of individual facilities. Also, grocery stores were eliminated as the possible target for this research, since, unlike big-box retailers, their supply chains are highly decentralized (e.g. use of vender-managed inventory (VMI) system), making data collection extremely difficult if not impossible. Also the experts stated that, in general, major retailers operate their shipping and routing schedules based on highly standardized supply chain management schemes, and thus retailers in different market may share the same characteristics in terms of TTG prediction is concerned.

One of the interviewees provided two important insights on the development of the model framework. First, the interviewee emphasized the importance of employee and facility size as the proxies for estimating the number of truck trips. His remark was consistent with many past studies reviewed in Chapter 2. Second; the most important variable that the interviewee emphasized was customer demand. According to him, sales forecasts have a direct impact on TTG since businesses usually adjust SCM strategies to meet the future customer demand. Figure 3-4 shows an example of how the sales volume forecasts affect the decision on delivery schedules. This company plans the level of production based on the sales volume forecasts for their chain of stores. The production level is the basis of planning for inputs such as raw materials. Finished goods are moved to the warehouses, and then are shipped to each store based on the delivery schedule so that the stores receive just the right (or projected) amount of merchandise.

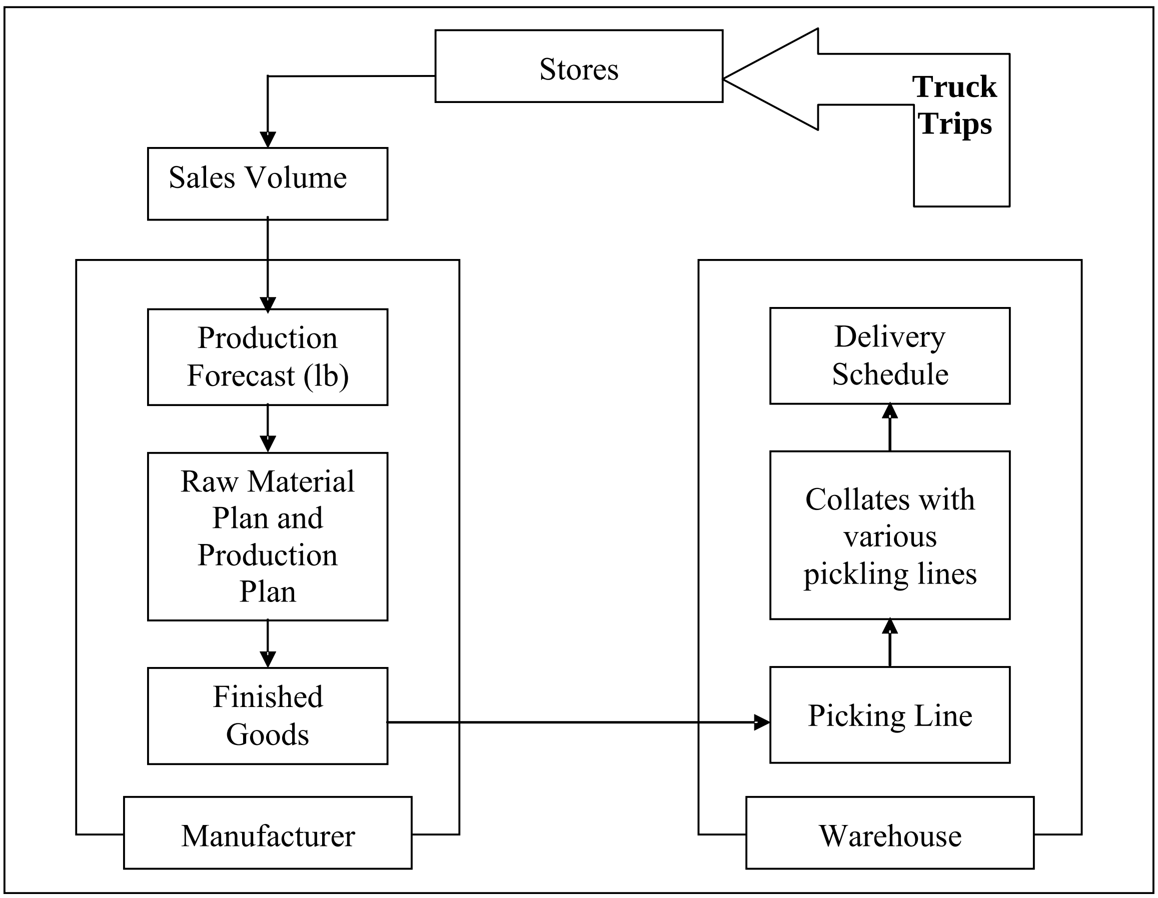


Figure 3-4 Demand Responsive Supply Chain System

The framework of TTG model for the retail businesses, depicted in Figure 3-5, is based on the following assumptions:

- TTG depends on business decisions which try to optimize the supply chain. That is, the number of truck trips is directly related to the activities at individual facilities where the strategies and actions may depend on the unique circumstances in order to maximize a facility's efficiency and profit by minimizing cost. Such strategies may translate to store-specific information, for example, number of employees, store size, sales volume forecast, and location of stores, the availability of loading and unloading facilities, and others physical site attributes.
- An important indicator of store performance is the sales volume that depends on consumer demand. Thus, socioeconomic characteristics of the market area of the store are related to the TTG.

The small box on the top of the figure implies that the starting point for the TTG model is the consumer demand for different types of goods. However, the inclusion of numerous commodities in a model is probably not feasible in most cases since the trade-off between the data requirement and the marginal improvement in the accuracy of the model predictions may not be favorable. Instead, it is assumed that there are four types of commodities: fast-moving and slow-moving goods in terms of the velocity of inventory turns and weigh-out and cube-out goods in terms of size and weight of the shipments.

The bottom of the middle box shows the variables that need to be considered in the TTG modeling process. These variables are classified as either long-term or short-term factors. The long-term factors include such variables as physical constraints of a facility and human resources. In this model, these variables are empirically tested in order to see if they are useful predictors for the TTG estimation. On the other hand, short-term factors are associated with the daily operation of the business. Variables such as replenishment schedule and sales volume are the most important variables. The variation of the sales volume over time will show the seasonal variations in business operation that may be related to the number of truck trips.

Three boxes on the bottom of the figure represent suppliers of commodities. As noted earlier, there are several different shipment patterns between a retail store and suppliers or DCs. Only three types of inbound shipments to a retail store are depicted.

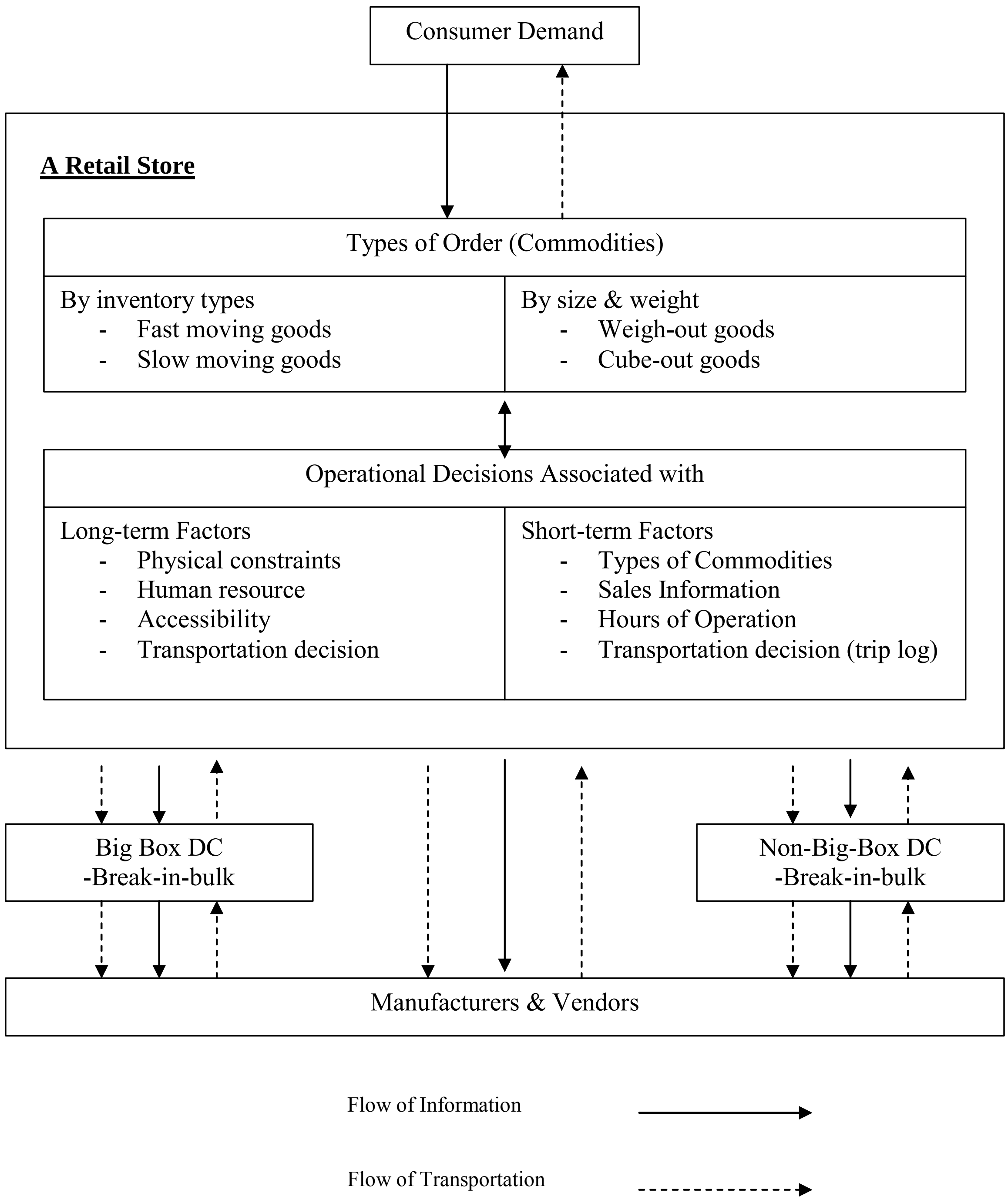


Figure 3-5 Conceptual Model of Truck Trip Generation

3.4 Chapter Summary

This chapter has provided the framework for TTG modeling that this study will employ. The discussion began with a simple model of the supply chain of a business. Then, the justifications for constructing the model at the disaggregate level were presented. Then the information obtained from the interviews of industry experts and their implications for the TTG model were discussed. Finally, a specific model for the retail sector, the target sector for this study, was developed and key assumptions were derived.

The next chapter discusses the strategies that were used to collect the data that are necessary to calibrate the model depicted in Figure 3-5 and conduct quantitative analyses.

4 . DATA COLLECTION STRATEGIES

This chapter discusses the strategies used to obtain the data needed to calibrate the TTG model for big-box retailers. The steps of data collection were carefully designed as an exploratory research process as explained in the following sections. First, available public data sources are reviewed in section 4.1. Then, sections 4.2 and 4.3 discuss the efforts that developed the contact list and the data wish list. Detailed data collection steps are presented in sections 4.4 and 4.5. The chapter concludes with a summary of the data collection efforts.

4.1 Search for Available Data Sources

To construct the model presented in Chapter 3, following data are required:

- The number of truck trips from the DC to stores (i.e. routing schedule or replenishment schedule)
- Types of Commodities by: inventory types (i.e. fast-moving vs. slow-moving), and by volume-to-weight ration (weight-out vs. cube-out).
- Store operation-related information
- Physical constraints (i.e. floor space, number of docks, number of doors, etc.)
- Number of employees
- Sales volume
- Consumer demand

Since commercial data sources are expensive and are often of questionable quality, the first step was to scour the public sources for these data. Publicly available data offer many advantages besides the cost. In most cases, they offer higher level of statistical reliability, or at a minimum better documentation, than commercial data. Usually, they are accompanied by the documentation on data collection strategies, statistical reliability, and potential problems, which is often not the case for commercial data sources. For some types of data, such as socioeconomic characteristics, public sector can be the main source. In contrast, existing public sources on freight data suffer from various problems. The website of the Bureau of Transportation Statistics (www.bts.gov) provides freight-related information such as rail waybill sample (in a summarized form), inland water freight movement, and the Commodity Flow Survey (CFS). The CFS is probably the most well-known data set. It includes the data, which are collected from a survey of shippers, on the flow of commodities by mode of transportation (BTS, 2005). It is conducted every five years, and the 2002 CFS is the most recent release. The commodities are classified using 5-digit Standard Classification of Transported Goods (SCTG) code. Value, tonnage, ton-miles, and mode of shipment are provided. Also, limited origin and destination information between states or selected metropolitan area is publicly released. However, the biggest weakness of the CFS is that metropolitan area is the most detailed geographical level available. Another disadvantage is that the CFS is a survey of shippers and does not report the data in terms of trips. This has been criticized in literature because the CFS-based models cannot capture the empty trips and trip chains (Fisher & Han, 2001; Holguin-Veras et. al. 2001).

Aggregate data regarding sales, facility size, and annual sales volume are available from the Census Bureau. The most well known survey is the Economic Census that is published every five years. The Economic Census provides information on “Nation's economy once every five years, from the national to the local level." (US Bureau of Census, 2005) Industry sectors are classified based on North American Industry Classification System (NAICS). Information on the number of establishments by employment size (in classes) is available at the six-digit NAICS level for each zip code. However, the facility-level information is not available.

The socioeconomic data are available from the Census of Population by the U.S. Bureau of Census. The data are released every 10 years. The most recent census was conducted in 2000 (U.S. Bureau of Census, 2005). The census contains various socioeconomic characteristics such as population (e.g. total, race, age, sex, household, family, etc.) and income (e.g. aggregate income, median household income, etc.). Twelve types of census files are available at the various levels of census geography[2] (U.S. Census, 2005).

Travel behavior data can be obtained from 2001 National Household Travel Survey (USDOT, 2004). The data are available for downloading as a spreadsheet or a raw data from the web site maintained by the Bureau of Transportation Statistics (BTS, 2005). The survey covers personal daily trips including long distance trips. Specifically, person and household trips are surveyed by trip purposes.

The data on the number of employees, sales volume and store size in square feet, were purchased from InfoUSA, which is a private firm that collects and sells a variety of business information (InfoUSA, 2005). There were three reasons for purchasing this data. First, the number of employees, sales, and store size are not publicly available. In addition, the research team experienced difficulty in collecting this data from the DCs and stores directly. There were some concerns about confidentiality and there were some occasions where other business departments held this data. Second, it offered a higher level of accuracy. A small sample of data was purchased from another source of business data, and it contained too many missing values for the store size, and was not usable. Moreover, the number of employees is categorized by range, making the data less useful. By contrast, the information for 88 percent of surveyed stores (See Table 5-2 in Chapter 5) was available from InfoUSA.

4.2 Creating Contact Information

Considering the amount of resources available and following the recommendation made during the preliminary interviews, it was decided that the data collection efforts would focus on the DCs and retail stores in the Midwestern states. A free website operated by Hoovers (www.hoovers.com) provided, for major businesses in the U.S., detailed information including: name, brief descriptions, and competitors of the business (APPENDIX A).

---

[2] The basic hierarchy of the census geography is: nation, regions, divisions, states, counties, census tracts, block groups, and blocks in order of size. The detailed information can be found at the U.S. Census Bureau web site.

Because of the time limitation, it was clearly impossible to cover the entire retail sector. Thus, our strategy targeted the sectors that sell limited types of commodities with high levels of homogeneity in size and weight. This strategy was used because a small number of commodity types would likely to mean a simpler supply chain and more importantly, less data collection. The homogeneity in size and weight would reduce the need for collecting the data on the mix of the merchandise sold, and also permit the use of the total sales (in dollars) as the proxy for the size and weight of goods sold. After considering these factors, footwear retailers, wholesale-type general retail chains, and home furnishing stores were selected as the main targets for the survey.

The next step was to create a sampling frame by generating a contact list for potential survey subjects. This research focused only on the retailers with a nationwide chain of stores because they provided an opportunity to collect a large amount of data from a single contact. It should be noted that small neighborhood-type stores may operate completely different supply chains, and thus the model developed here needs to be applied with a caution. In addition, it was assumed, based on the preliminary interviews, that major retailers operated shipping and routing schedules based on a highly standardized supply chain management scheme relative to smaller businesses, therefore the findings from a study of larger retailers may be transferable to other major businesses with similar characteristics.

The addresses of 150 retailers were obtained relatively easily through the business websites on the Internet. However, the contact information for potential data sources, that is, the personnel in logistics or supply chain management position within each company, was difficult to obtain. Thus, the membership directories of professional organizations in the field of logistics, retail, and supply chain were used to identify several contacts within each firm and to obtain the phone numbers and e-mail addresses for those contacts.

4.3 Data Wish List

In parallel with the creation of the contact list, the research team generated a data wish list (APPENDIX B). The purpose of the wish list was to communicate, to the potential respondents, the types of data needed for the analysis and also suggest possible sources where the information could be found. Once the businesses agreed to participate in the survey, the wish list was sent along with instructions for delivering the data. Compared to using a survey instrument, this approach provided the contacts more flexibility in terms of the format and the means to deliver the data. The drawback was that it required a considerable amount of follow-up with phone calls or other means to go over the list to ensure that the contacts had a clear understanding of the data being sought and to make necessary arrangements to obtain them

The language in the wish list had to be carefully developed and refined in order to clearly convey what the research team needed. For example, the most critical material was a "travel diary" between a DC and retail stores. Although travel diaries are common tools used by transportation planners to collect travel information, it was found that logistics professionals normally referred to them as "route schedules" or "replenishment schedule".

4.4 Supplemental Survey

In addition to the main data collection effort using the contact directory and the data wish list, which relied mostly on phone calls and e-mails to the DC managers, the second approach – store visits and phone calls to individual stores – was conducted. As described in the following sections, two approaches were used in a manner that complements each other.

4.5 The Survey

4.5.1 Distributing Questionnaire: Contact List and Data Wish List Approach

Once the contact list and the data wish list were created, several rounds of phone calls were made to each contact. Each company was contacted 3-5 times. About 50% (37 in 75) of contacted companies agreed to review the data wish list. As seen in Table 4-1, for the main survey, the data were obtained from only three different retail companies (Furniture Chain A, Shoe Chain A, and apparel chain A). The response rate was extremely low at only 4%.

Table 4-1 Survey Results

Strategy	Chains	SIC*	Response Rate	Number of Stores	Advantage	Disadvantage
Distributing Survey Questionnaire to Distribution	Furniture A	5712	4% (3/75)	76	* Detailed replenishment	* Extremely low response rate * Time consuming * No consistent information
	Shoe A	5661		259		
	Apparel A	5632		n/a		
	Subtotal			**335**		
Visiting Stores	Furniture B	2511	100% (4/4)	4	* Cooperative	* No detail information * Time consuming
	Furniture C	5719	100% (6/6)	6		
	Furniture D	2512	100% (6/6)	12		
	Subtotal		100% (16/16)	**22**		
Phone Survey (Individual Stores)	Furniture E	5719	26.1% (13/36)	13	* Somewhat responsive * Low cost	* No detail information
	Shoe B	5661	67.5% (27/40)	27		
	Shoe C	5661	50% (10/20)	10		
	Shoe D	3149	79.2% (19/24)	19		
	Subtotal		59% (69/117)	**72**		
Total				**426**		

* The definitions of 4-digit SIC

5712 - Furniture Stores
5719 - Micellaneous Home Furnishings Stores
2511 - Wood Househod Furniture, except upholstered
2512 - Wood Household Furniture, upholster
5661 - Shoe Stores
3149 - Footwear, Except Rubber, Not Elsewhere Classified

On the other hand, the advantage of this strategy is that highly detailed information can be obtained from the interviewees. For example, the research team visited the DC of Furniture Chain A. The DC covers 76 stores in 18 states in the Midwestern and Eastern states. In addition to routing information for a typical week, a day-long meeting provided the research team with valuable insights on the operations of interviewee's company and other DCs in a similar sector. Shoe Chain A provided the routing information for the entire year of 2004 for 259 stores in 23 states.

Despite these cases of success, numerous problems prevented the research team from pursuing this strategy further. First, while the contact list reduced the problem to some extent, it was still problematic to reach the right person within each company. Second, much of the data requested in the wish list were considered confidential by the businesses. Consequently, the decision for providing the data had to be made at a high level even when the data were available, causing delay and an extremely low response rate. Third, in many cases, the businesses had to be provided with a certain incentive or motivation to participate in the study. One possible incentive was to provide the study results or summarized data to the participants so that they could use it as the benchmark. However, this obviously did not entice many businesses to participate. Fourth, detailed information on the number of employees and sales volumes by stores was not available.

4.5.2 Store Visit and Phone Survey of Individual Stores

To supplement the data obtained via surveys, the research team also contacted the stores directly by visiting or calling them. The objectives of this approach were to complement the data obtained by the main survey, and to also verify if the stores in the same sector received the similar number of weekly deliveries with other factors being equal. In addition, the data points were to be used to validate the performance of the TTG model based on the detailed information collected from the main survey. Thus, the competitors of Furniture Chain A and Shoe Chain A were identified from the websites of Hoovers (www.hoovers.com) and Investors Words (www.investorwords,com).

A total of 16 stores in the area were visited. While the response rate of 100 percent was encouraging, it was a resource intensive method (e.g. time and cost). By contrast, the phone survey proved to be quite efficient. A total of 117 stores were called. The overall response rate was 59 percent (69 stores). It was especially encouraging that most of those respondents provided the most cortical information, the number of truck deliveries per week, if not more detailed store-specific data.

4.6 Chapter Summary

This chapter discussed the data collection efforts. Based on the recommendations given by the experts, the research team targeted the DCs of national retail chains. Despite extremely low response rate, two data sets, from Furniture Chain A and Shoe Chain A, provided detailed information on their routing schedules and other variables. The resource requirement and low response rate of the approach prompted contacting individual stores directly by visiting or calling them. With this approach, the response rates were very high; however, only the number of delivery tripsfor each store was available. The next chapter will describe in detail the collected data and the database building process.

5 . DATABASE CREATION AND DESCRIPTION

Four data sets were used in a complementary manner to create the databases for statistical analyses. The first data set is from the survey described in the previous chapter. Although the level of detail and the quality of information varied depending on the survey method and the respondent, trip information on a total of 426 stores was collected. The InfoUSA data provided store-specific information such as the number of employees, sales, and size for the stores in the survey data. The third dataset was the 2000 Census. The census block group level population and income information from Census Summary File 3 (SF 3) was used to compile socioeconomic characteristics of the market areas for the stores in the survey data. The final data were from ESRI Street Map U.S.A. (ESRI, 2004), which provided the address-level map used to geocode the stores, enabling the spatial matching of the SF3 data with the market area of each store. All data were merged in the geographic information systems (GIS).

5.1 Description of Data Sets

5.1.1 Survey Data

As shown in Table 4-1, five furniture chains, four shoe chains, and one apparel chain provided trip information. The data for Furniture Chain A contained the information on 76 stores in 18 states. The dataset included the number of deliveries (one or two per week) to the stores, the number of pallets per delivery, routing schedules for a week of February 11, 2005, store addresses, store location characteristics (off-mall or mall), and store types (combo, conventional, or outlet). Combo stores sell both regular and children's furniture while conventional stores only sell regular furniture. The outlet stores handle returned and out-dated furniture. The DC manager revealed that the company is shifting the stores from the malls to off-mall locations, and those new stores are mostly the combo type.

Shoe Chain A provided detailed daily routing schedules for 259 stores in 23 states for the entire year of 2004. The data contained the number of deliveries per week to each store, the number of cartons per delivery, and the routing schedules. Interestingly, all the stores received one delivery a week without exception. In other words, there was no variation in the number of trips. Thus, the data could not be used for the calibration of TTG models. Only the data on the number of cartons delivered was used for exploratory and regression analyses, discussed in Chapters 6 and 7.

Data for an additional 94 stores were collected by visiting stores or making phone calls. In spite of high response rates, only the number of weekly deliveries was obtained for each store, thus limiting the usefulness of the data to the verification of the models. Furthermore, only the Furniture Chain E shows any variation (one to two deliveries) in the number of deliveries per week. In Chapter 7, the dataset for Furniture Chain E was used to verify the transferability of TTG models.

5.1.2 Socioeconomic Data

As stated earlier, one of the assumptions of the study is that the sales volume of a store, which reflects the underlying consumer demand within the market shed, is one of the determinants of the store's TTG rate. Thus, the variables that capture the underlying demand should be examined as the potential independent variables in the TTG model.

The census block group level information in the 2000 Census Summary File 3 (SF3), available from the U.S. Census Bureau web site (www.census.gov), was used to compile the socioeconomic data. SF 3 includes the population by different socioeconomic characteristics such as age, household types, income class, as well as aggregate income, which is often used in market research to measure consumer demand (McCarthy, 2001), and median household income. Table 5-1 presents the 37 socioeconomic variables considered in the TTG model development. With a few exceptions, all the variables are taken directly from the SF3 dataset without adjustments. Modifications include: youth, workforce, and retire were regrouped based on ages, and also low income, mid income, and high income were grouped based on the number of people by income groups.

Table 5-1 List of Socioeconomic Variables

Variables	Description
pop2000	2000 population
pop00_sqmi	Population density
white	White population
black	African American populaiton
ameri_es	American Eskimo population
asian	Asian population
hawn_pi	Hawaiian and Pacific Islander Population
other	Other races
mult_race	Multiple-race
hispanic	Hispanic
males	Male
females	Females
med_age	Median age
med_age_m	Median age of female
med_age_f	Median age of male
households	Number of households
ave_hh_sz	Average household size
hsehld_1_m	Number of one-person male household
hsehld_1_f	Number of one-person female househod
marhh_chd	Number of family households with a married couple and related childeren under 18 years
marhh_no_c	Number of family households with a married couple and no related childeren under 18 years
mhh_child	Number of family households with a man and related children under 18 years but no wife
fhh_child	Number of family households with a woman nad related childeren under 18 years but no husband.
families	Number of families
ave_fam_sz	Average family size
hse_units	Number of housing units
vacant	Number of housing units that are vacant
owner_occ	Number of housing units that are occupied by the owner
renter_occ	Number of housing units that are occupied by renters
mincome	Median Income ($)
aincome	Aggregate income ($)
youth	Number of people under 21 years old
workforce	Number of people between 21 and 64 years old
retire	Number of people over 65 years old
lowincome	Number of households whose household income is less than $35,000

5.2 Database Building Process

This section describes how the datasets from different sources were merged to form a single dataset. Figure 5-1 is a graphical representation of the database building process that consists of three steps. These steps were performed separately for each store chains.

5.2.1 Geocoding

In this step, called "geocoding[3]", each store was assigned geographical coordinates that correspond to its street address. ESRI Street Map U.S.A was used as the base map for this process. Of 426 stores, 97.9 percent, or 417 stores, were successfully geocoded. Specifically, all 76 stores of Furniture Chain A, and 247 out of a total of 259 stores belonging to Shoe Chain A were correctly geocoded.

5.2.2 Buffer Creation

Second task is to create a buffer for each store. The buffer is created to represent the market shed, or the area within which the store attracts customers. Since each driver uses a "cost-minimizing approach to individual travel behavior and the generation of multiple purpose trips (Pipkin, 1995)," if other conditions being equal, a driver may choose the store within a shortest distance. That should lead to a natural formation of a market area (more commonly known as "market shed" in transportation field) for each store. The market areas were estimated in ArcGIS by creating a circle, having a radius equal to a certain travel distance, with the geocoded coordinate of each store at the center. Travel distance was determined based on the lengths of shopping trips reported in the 2001 National Household Travel Survey (NHTS) data (U.S. DOT, 2004).

---

[3] "Geocoding is the process of assigning a location, usually in the form of coordinate values, to an address by comparing the descriptive location elements in the address to those present in the reference material. Addresses come in many forms, ranging from the common address format of house number followed by the street name and succeeding information to other location descriptions such as postal zone or census tract" (Crosier, 2004)

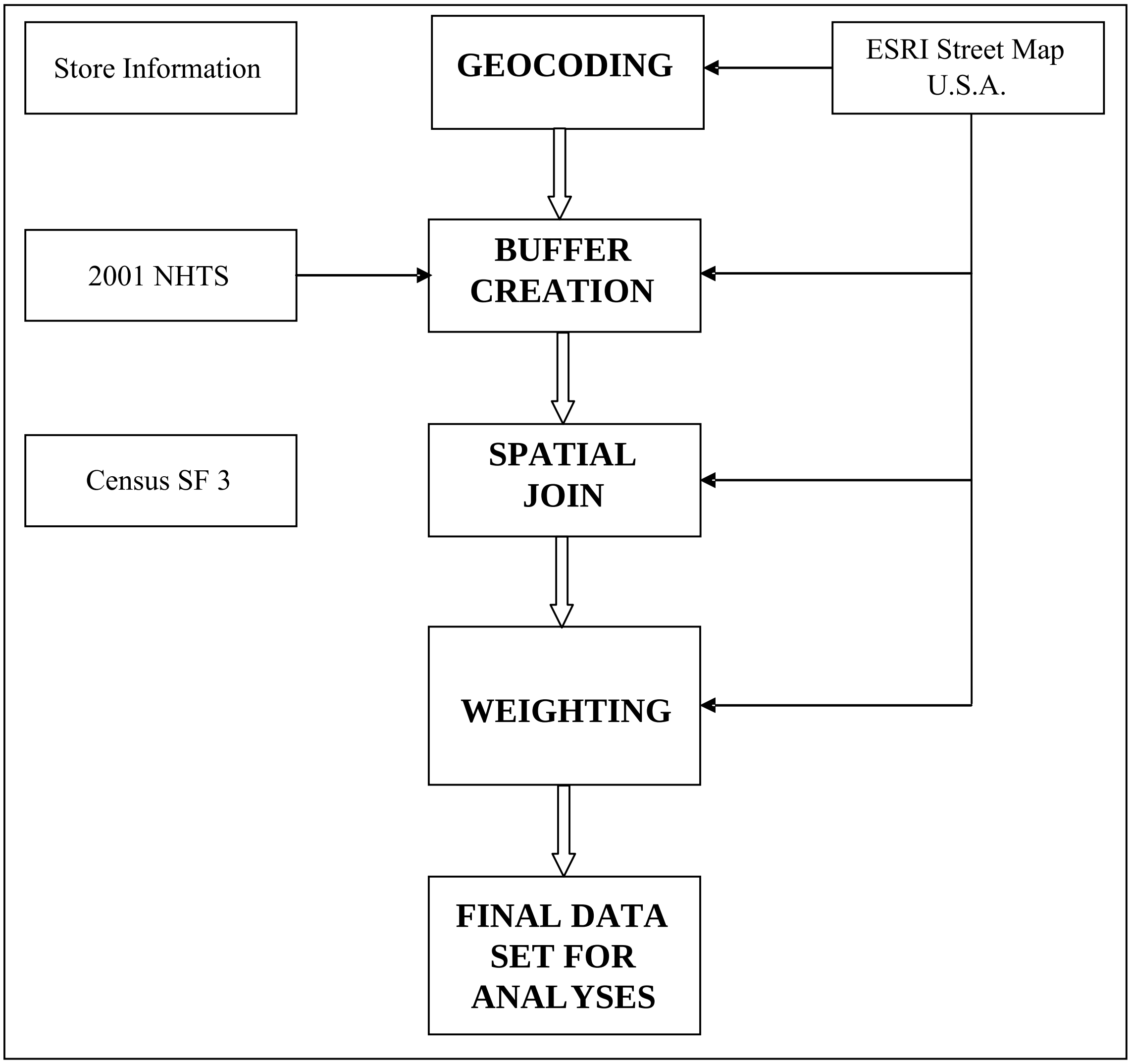


Figure 5-1 Flow Chart of Database Building Process

The NHTS data summarizes passenger travel characteristics by various trip purposes and demographic categories. As in the Census, NHTS provides travel characteristics by location such as urban, suburban and rural areas. The 75th percentile of reported shopping trip lengths in the urban areas, after removing outliers, was calculated to be 6 miles. Thus, this value was used to create a buffer, with a 6 mile radius, around each store in ArcGIS.

5.2.3 Spatial Join

In this step, buffers were superimposed on the GIS layer for the census block groups. The block groups that either intersected or were within each buffer were identified and spatially joined with the buffer. Then, the socioeconomic information from the 2000 Census SF3 for each block group was attributed to each buffer using the block group ID numbers.

5.2.4 Weighting

At the end of the procedure described in the last section, the dataset included, for each store, store characteristics, socioeconomic information of market areas, and geographical coordinates. Each buffer contained a number of block groups, some right next to the store and others far away. It was reasonable to assume that the residents of the nearby block groups have greater impact on the merchandise sales at the store than those living far away simply because the latter should be less likely to travel to the store.

In order to quantify the effect of the distance between the residential location and the store, following four weighting schemes were applied to the socioeconomic data:

- Weight 1: No weight
- Weight 2: Inverse of distance
- Weight 3: Inverse of distance squared
- Weight 4: Inverse of distance cubed

The aerial distance between the store and the centroid of each block group, calculated in ArcGIS, was determined using the coordinates of both points.

5.2.5 Missing Values

Missing values found in the data sets for this research could not be replaced with any of commonly used techniques such as average value, past information, or historical trends.(Buchheit, 2002). Thus, missing variables were simply removed from the data set. Table 5-2 summarizes the usable values after deleting missing values. The second column shows the number of data points for which TTG information was obtained. The columns 3 to 6 present the number of usable data points (and percentages if there was any missing values) for each of four types of datasets that were used to create the final dataset. Nine rows with chain names represent data sets from the surveyed stores. The last column shows the percentage of usable observations after removing missing values.

Table 5-2 Usable Observations

	Truck Trips	Geocoded Observations	Number of Deliveries	Delivery Units	InfoUSA	Usable Observations
Furniture Chain A	76	76	69 (91%)	76	64 (84%)	58 (76%)
Furniture Chain B	4	4	4	NA	2 (50%)	2 (50%)
Furniture Chain C	6	6	6	NA	4 (67%)	4 (67%)
Furniture Chain D	12	12	12	NA	4 (33%)	4 (33%)
Furniture Chain E	13	13	13	NA	11 (85%)	11 (85%)
Shoe Chain A	259	242 (97%)	221 (85%)	242	240 (93%)	207 (80%)
Shoe Chain B	27	27	27	NA	24 (89%)	24 (89%)
Shoe Chain C	10	10	10	NA	9 (90%)	9 (90%)
Shoe Chain D	19	19	19	NA	16 (84%)	16 (84%)
Total	426	417 (98%)	381 (89%)	308 (100%)	374 (88%)	335 (79%)

(xx %) - percent of original data points that are usable

5.3 Chapter Summary

This chapter presented the process of dataset development. Four datasets were integrated using the ArcGIS to create the final database. Four different weights were applied to socioeconomic variables to account for the effects of distance to the store, producing a total of four data sets. The detailed analyses, using this dataset, will be discussed in Chapters 6 and 7.

6 . EXPLORATORY DATA ANALYSIS

The goal of this chapter is to describe surveyed data sets. The chapter begins with a brief discussion of the findings from the surveyed datasets. Then, section 6.2 examines various dimensions of the Furniture Chain A dataset. Then, the Shoe Chain A dataset is explored in section 6.3. The insights obtained from the exploratory analyses facilitated the construction of the TTG models.

6.1 Findings from the Surveyed Data

Table 6-1 summarizes the number of delivery trips to the stores for each of nine retail chains. Several interesting trends were revealed. First, all surveyed retailers have a highly standardized routing schedule, supporting one of the basic assumptions of this study. Except for Furniture Chains A and E, all the stores of other seven chains receive the same number of deliveries throughout the year. There is no variation in the replenishment frequencies among the stores of each company. For Furniture Chains A and E, some stores receive only one delivery per week while the others receive two deliveries per week. Such standardization is made possible by consolidating and managing all the replenishments through the company owned DCs, which is similar to the system depicted in Figure 3-3. Second, an evidence of the shift from push to pull-logistics was found in Furniture Chain C. All the stores of this chain receive daily parcel shipments on the weekdays. The research team was informed that until the spring of 2005, Furniture Chain C used the replenishment schedule similar to other chains (two weekly shipments by semi-tractor trailers). The purpose of the change was to quickly respond to changing consumer demands. Third, the data indicate that Standard Industrial Classification (SIC) is not an effective scheme to categorize the businesses in terms of their TTG characteristics. Especially, this is the case for the furniture stores. Only two of five furniture chains share the same SIC numbers. It should also be noted that Shoe Chain D is categorized as manufacturer although it is generally recognized as a discount shoe store chain just like other three chains in the dataset.

6.2 Furniture Chain A

6.2.1 Store Locations

The distribution center for Furniture Chain A covers 76 stores in 18 states in Midwestern and Eastern parts of the U.S. Illinois has 17 stores, followed by Ohio (16 stores), Michigan (8 stores), and Indiana (7 stores). Stores receive one or two weekly deliveries throughout the year. For this chain, the most important information for tracking the performance of each store is the number of pallets delivered per week. In other words, a pallet is the basic unit of measurement for recording the volume of sales and approximate dollar value of each shipment, and also determining the number of trucks required for a delivery.

Table 6-1 The Number of Weekly Deliveries per Store

	Chains	Number of Deliveries	Truck Types	Coverage per routing	SIC
Furniture	A	1 or 2	Semi	2-4 stores	5712
	B	1	Semi	n/a	2511
	C	5	UPS	n/a	5719
	D	1	Semi	n/a	2512
	E	1 or 2	Semi	n/a	5719
Shoe	A	1	Semi	3-5 stores	5611
	B	2	Semi	n/a	5611
	C	1	Semi	n/a	5611
	D	1	Semi	n/a	3149

According to the interviewee, the average dollar value of furniture per pallet is $515. Up to 28 pallets can be shipped in a 53-foot container. Most trucks leave the DC when they are approximately full. Thus, it is possible to approximate the number of truck trips originating from the distribution center by dividing the total number of pallets by the load capacity of a truck. However, with this data, it is not possible to approximate the number of truck trips to each store, since each truck can cover 2-3 stores in a single trip, or "a route". In this dataset, the maximum number of pallets that a store received is 18.

The data source revealed during the data collection that the company categorizes each store by type (conventional, combo, and outlet) and location (mall and off-mall). He/she also revealed that the company designed the replenishment schedule partly based on those characteristics. The stores with characteristics such as Combo and/or off-mall tend to receive two replenishments per week. This is partly driven by the sales volume, which tends to be higher for combo and/or off-mall stores, and also the company's overall strategy to target the market segment that is attracted by those types of stores.

6.2.2 Relationship between Delivery Frequency, and Store Types and Locations

The data on the number of pallets delivered per week were available for 69 stores. As discussed previously, the trip frequency for this chain is highly standardized. Fifty one stores, or 74 percent, receive one delivery and the remainder, 18 stores (26 percent), receive two deliveries. Furniture Chain A has three types of stores (combo, conventional, and outlet stores) and two location characteristics (off-mall and mall).

Figure 6-1compares the replenishment frequencies by location characteristics of the stores. While the majority of the stores receive one delivery a week, the graph indicates that off-mall stores are more likely to receive two weekly deliveries. Figure 6-2 compares the frequencies by store types. Compare to other two types, 9 of 10 combo stores receive two deliveries per week. Figure 6-3 compares replenishment frequencies by both store type and location. The figure reveals that seven out of eight stores that are both combo type and off-mall receive two weekly deliveries. This information will be used in the model building, discussed in Chapter 7.

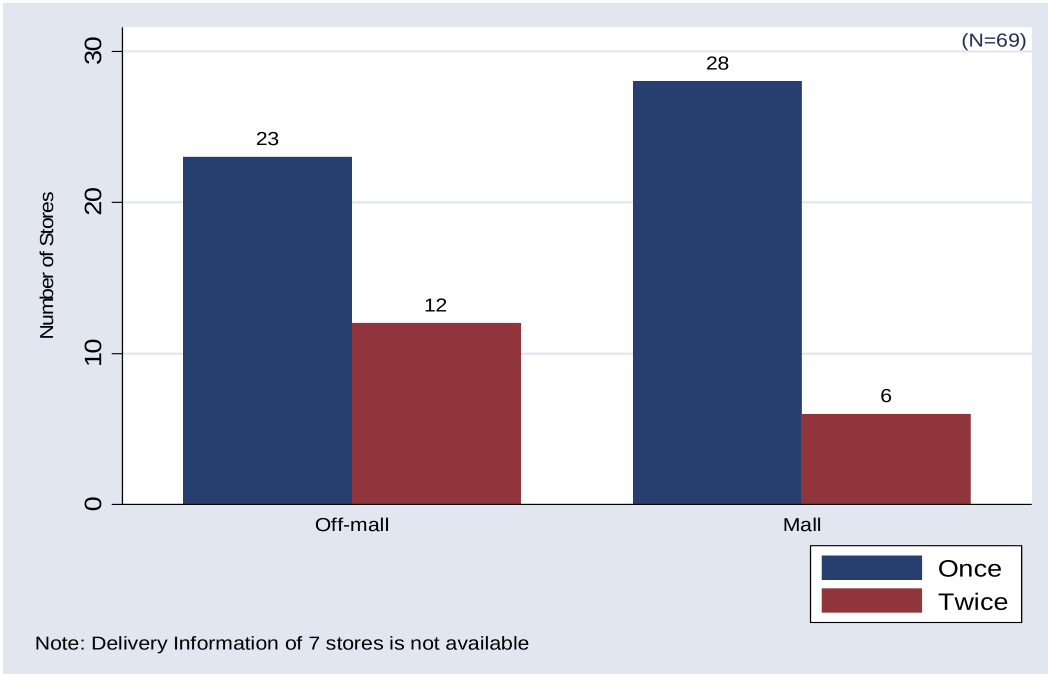


Figure 6-1 Delivery Frequencies by Location Types - Furniture Chain A

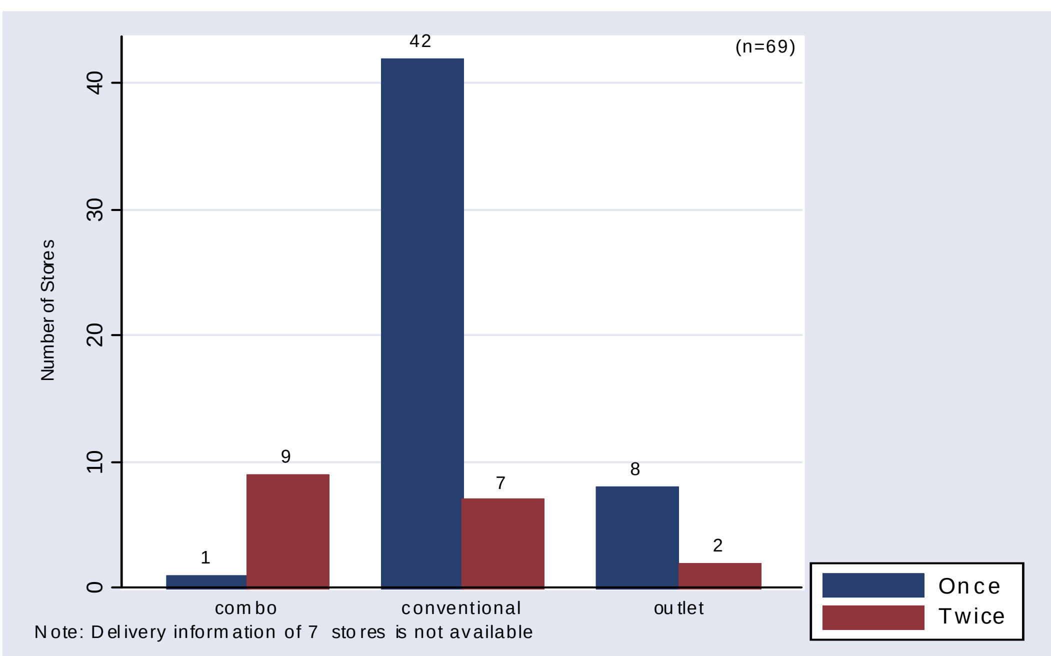


Figure 6-2 Delivery Frequency by Store Type - Furniture Chain A

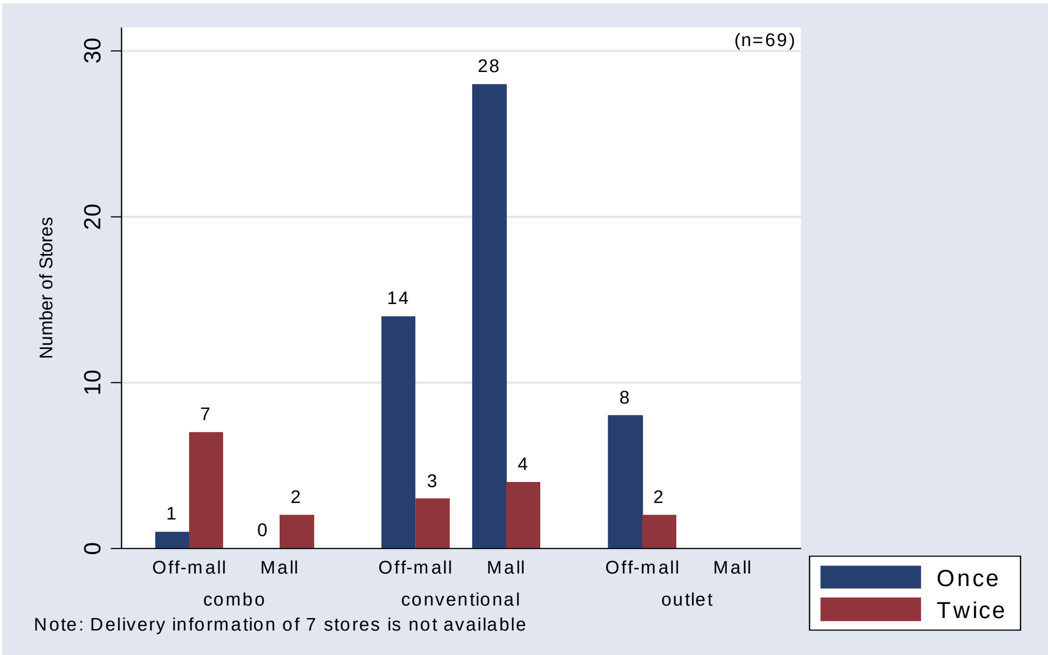


Figure 6-3 Delivery Frequency by Location and Store Type - Furniture Chain A

6.2.3 Relationship between Weekly Pallets, Delivery Frequency, Store Types and Locations

This section considers an additional dimension by analyzing the number of pallets delivered. Table 6-2 summarizes the data on the number of pallets delivered. The overall average is 9.01 pallets per delivery, with a minimum of 4 pallets and a maximum of 16 pallets. Since 18 stores receive two weekly deliveries, the average number of pallets per week is also computed. The average number of pallets per week is 11.06 with minimum of 6 pallets and maximum of 32 pallets.

Table 6-2 Summary of Pallets Delivered - Furniture Chain A

	Mean	S.D.	Min	Max
Pallets per delivery	9.01	2.73	4	16
Pallets per week	11.06	4.64	6	32

When store type and location type are considered, complex relationships can be observed. Figure 6-4 shows the number of pallets delivered per week by store types and location characteristics. The numbers 1 and 2 on the x-axis represent delivery frequencies. In the figure, CO, CV, O stand for combo stores, conventional stores, and outlet stores, respectively. According to the graph on the upper left side, off-mall combo stores receive fewer pallets per week than the stores in malls. The conventional stores, shown in the upper right corner, in off-mall stores seem to receive more pallets per week than their mall-based counterparts. As far as the trip frequencies are concerned, it is very clear that off-mall based combo stores are highly likely to be two-delivery stores. However, in terms of the number of pallets per week, the relationship is somewhat unclear.

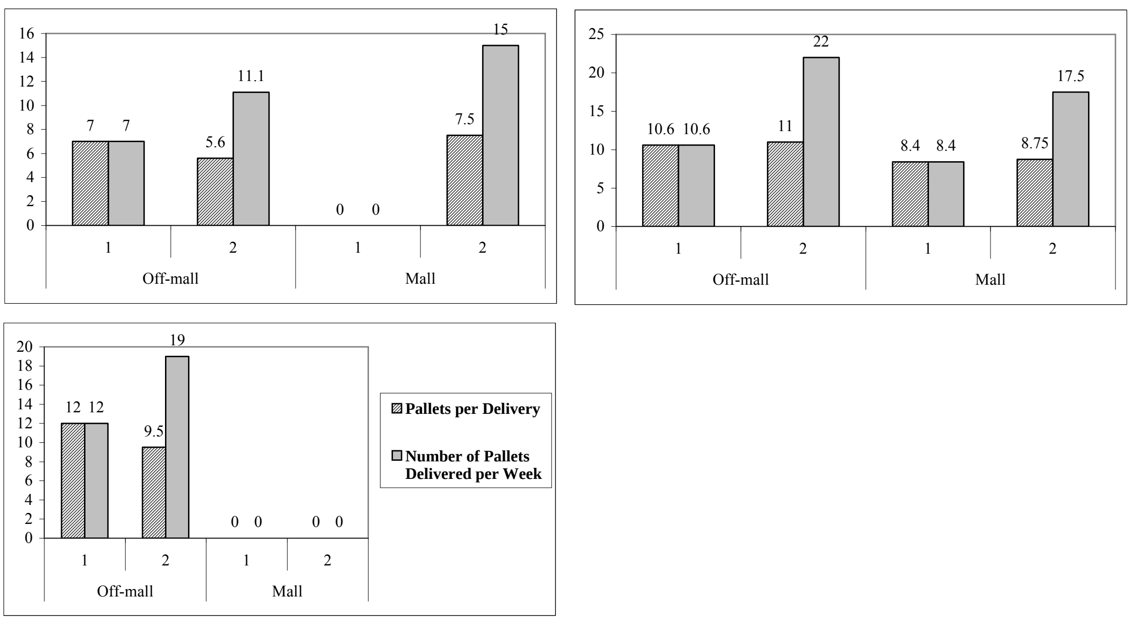


Figure 6-4 Delivery Frequency and Pallets Delivered - Furniture Chain A

6.3 Shoe Chain A

Shoe Chain A has only one DC that covers 259 stores in 23 states mostly in the middle part of the U.S. Twenty three stores are located in Illinois, followed by Indiana (19 stores), Texas (18 stores), and Missouri (16 stores).

The most interesting information in this data set is the variation in the number of cartons per delivery. A carton is the unit of shipment that contains between 6 pairs to 15 pairs of shoes depending on the size of the shoes. Thus, the number of cartons is the important information for the company to track the trend of consumer demand. According to the contact person at Shoe Store A, the average weight per carton is 22 pounds and the average weight per trailer load is 26,400 pounds. This equates to about 1,200 cartons per truckload.

On average, the stores receive 286 cartons per delivery, as shown in Table 6-3. The data show broad range as indicated by the fact that maximum number of carton is about 4.5 times as many as the minimum. However, the standard deviation is reasonably small relative to the mean. This implies the existence of extreme values. Additional analysis that divided the dataset into new (i.e. opened in 2004) and existing stores showed that the new stores often received extremely high number of cartons. This is due to the special pre-opening deliveries that were made to stock the new stores. For the TTG analysis, these pre-opening trips should be excluded since they are only temporary phenomena.

Table 6-3 Average Number of Cartons per Delivery per Store - Shoe Chain A

Mean	S.D.	Min	Max
286.67	79.55	151.86	727

Figure 6-5 displays the seasonal variation in the number of cartons shipped from the DC. Each of the twelve data points in the graph represents a month (from January to December). Interestingly, there seem to be three peak seasons in a year, coinciding with Easter, back-to-school, and Christmas. The graph indicates that the peak of replenishments, to meet the demand for those three occasions, occurred one to two months earlier. Thus, a very small amount of cartons were actually shipped in December, because the peak, which occurred in November. As mentioned earlier, all the stores receive exactly one delivery per week all year round. With the existence of seasonality in demand, as shown in the figure, the adjustments were made in terms of the load factor when the trucks leave the DC, and/or the number of trucks used. This calls for a rather sophisticated planning.

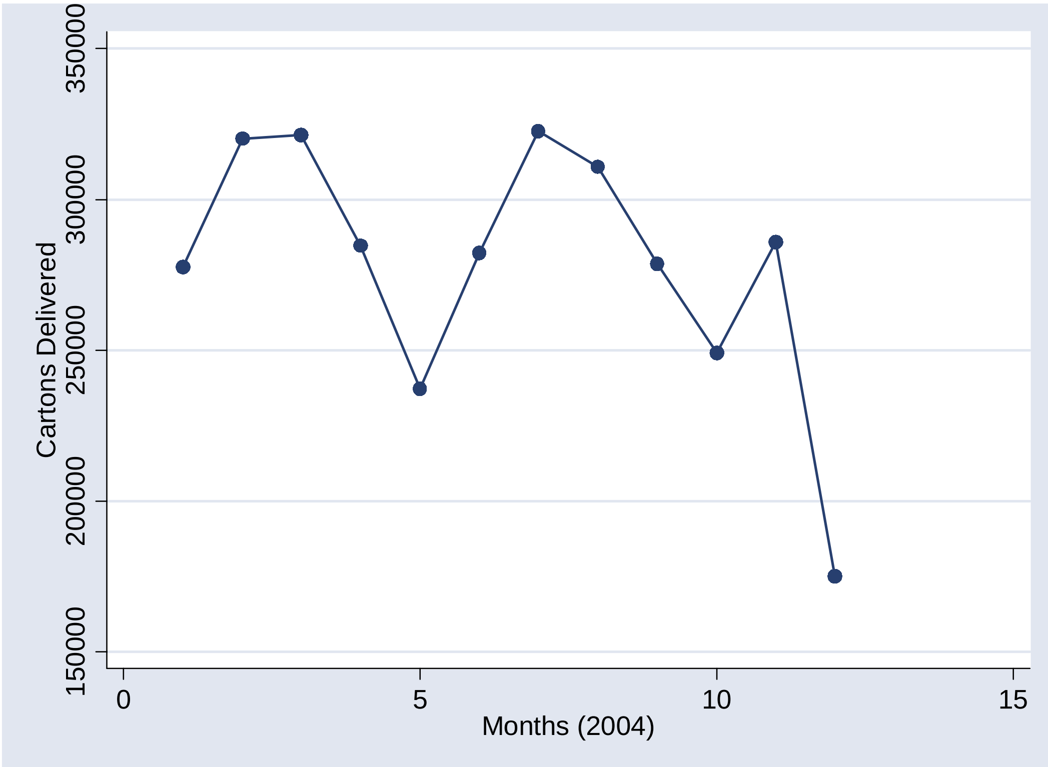


Figure 6-5 Total Number of Cartons Shipped from Distribution Center by Month - Shoe Chain A

7 . MODEL ESTIAMTION AND RESULTS

7.1 The Roadmap to Model Building

In this chapter, two distinct groups of models are developed. The first group of models, formulated by the multiple regression technique, examines the relationships between delivery units (e.g. pallets) and store and socioeconomic information. The second groups of models attempt to predict the delivery frequency, or delivery trips by trucks, using binary logit regression using the same variables. Due to the data limitation, the second type of models was developed only for Furniture Chain A.

The independent variables examined in this study can be classified into two classes. The first class consists of store-specific information. For Furniture Chain A, variables such as the number of employees, annual sales volume, store size in square feet, location characteristics, and store types were obtained. However, only the first two were available for Shoe Chain A. Detailed descriptions of the store-specific variables are given in Table 7-1. Two store size classes (*size 1* and *size 2*), two location characteristics (*mall1* and *mall2*), and three store types (*stchar1, stchar2, and stchar3*) were coded as dummy variables.

Table 7-1 Store-Specific Information

Variables	Descriptions
id	Store identification number
employee	Number of employees
sales	Annual sales volume ($)
size1	Store size in square feet: 0-9,999
size2	Store size in square feet: 10,000-39,999
mall1*	Store Location: Off-Mall-based stores
mall2*	Store Location: Mall-based stores
stchar1**	Store type: Combo
stchar2**	Store type: Conventional
stchar3**	Store type: Outlet

* Available only for Furniture Chain A and E

** Available only for Furniture Chain A

The second class of independent variables contains socioeconomic characteristics of the market area for each store that consist of the variables related to the population and incomes of the residents. The development of the dataset is discussed in Chapter 5.

The models are developed in a four-step procedure, displayed in Figure 7-1. The first step builds a model using only the store-specific information. Then, the models that include only the socioeconomic variables are constructed in the second step. As discussed in Chapter 5, different weighting schemes were used to generate a total of four separate socioeconomic datasets. Thus, four separate models were developed in this step. Also, variable selection processes were applied in this step. In the third step, the independent variables from the first and the second steps were considered at the same time to build the best model for each weight type. Finally, those four competing models were compared to identify the best model. Both multiple regression and binary logit model analyses for this research adhered to this procedure.

7.1.1 Multiple Regression Models for the Amount of Merchandise Delivered

As mentioned earlier, for the first group of models, multiple regression analysis was conducted to examine the relationships between delivery units and store and socioeconomic information. There are several standard assumptions associated with the multiple regression analysis. Most of these assumptions are required when applying the multiple regression to a sampled data set from a population. Since each of the datasets used in this study covers the entire population (i.e. all the stores covered by the surveyed distribution center are included), not randomly sampled stores, requirements such as normality and homogeneity of error terms does not apply. Multiple regression was used simply to find the best fit model. Essentially, the multiple regression was used solely as a line-fitting technique, not to perform statistical inference.

Thus, it does not make sense to seek the fulfillment of all of the standard regression assumptions. Instead, this research tried to avoid one problem that is still relevant when dealing with the population data, multicollinearity. It occurs when two or more independent variables in a model are highly correlated. When multicollineraity is present, the interpretation of the partial coefficients may not be valid.

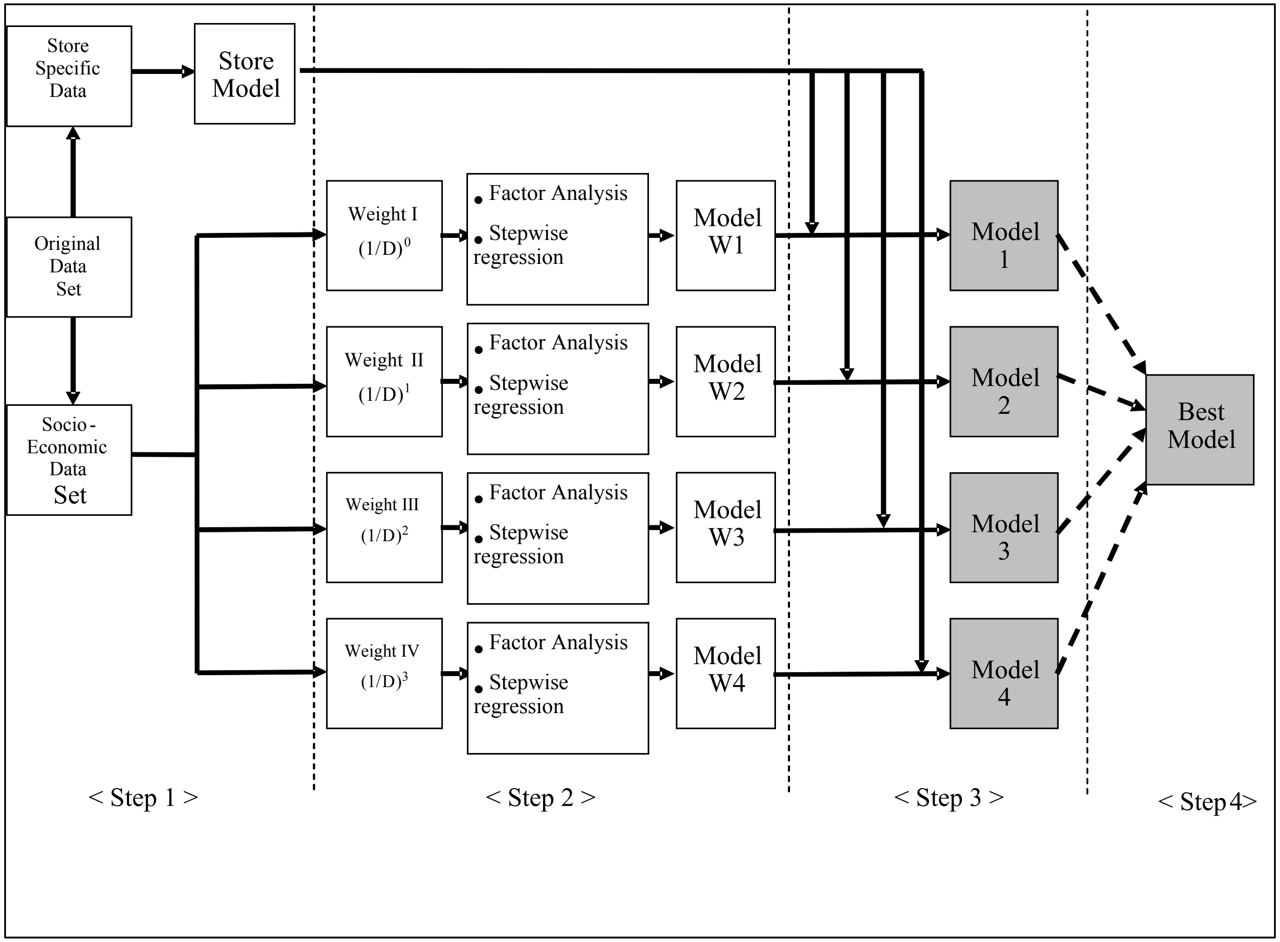


Figure 7-1 Model Building Process

The presence of multicollinearity was diagnosed with variable inflation factor (VIF)[4], Pearson correlation between independent variables, and the reasonableness of the coefficient estimates. For both store-specific and socioeconomic variables, various transformations were applied to address the multicolinearity problem and also improve the fit of the model.

R-squared measures the goodness of fit of a regression model. It is often interpreted as the proportion of the total variation in the dependent variable accounted for by the independent variables. The R-squared is a "nondecreasing" function of the number of independent variables (Gujarati, 2003)." In other words, R-squared keeps increasing as more independent variables are added to the model. Thus, when two models with different numbers of independent variables are being examined, it is difficult to compare the R-squared values. Alternatively, adjusted R-squares, which account for the number of independent variables, were used.

7.1.1.1 Variable Selection Procedures

The datasets included 37 socioeconomic variables. With this many variables to consider, a systematic strategy for selecting the independent variables was needed to streamline the process. Unfortunately, there is no universally accepted approach to determine how many and what variables should be included or removed from the model (Gujarati, 2003). Model building is a mix of science and art that inevitably involves researchers' subjective judgments. Nevertheless, there are several tools that can aid researchers to build the models with desirable traits. In order to select the set of variables that offer a high level of fit while avoiding multicollinearity, stepwise regression and factor analysis were employed to aid the process. Those two techniques provided a starting point for the variable selection procedure that involved a lengthy trial and error process before the final models are identified. A brief description for each of those two techniques is given below.

---

[4] . VIF is a measure of identifying the multicollinearity of independent variables. When VIF is over 10 the variables should be further investigated (http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm) .

7.1.1.2 Stepwise Regression

As a variable selection method, the most useful method is probably “the best subset regressions (Sen and Srivastava, 1990)” that considers all possible combinations of independent variables. However, the availability of resource limited the application of the method for this research. As an alternative, stepwise regression was used. Three techniques of stepwise regression are backward elimination, forward selection and stepwise procedure. The backward elimination starts with all independent variables (37 variables for this research) in the model and eliminates the less significant variables one at a time. The reverse of the backward elimination is the forward selection. A combination of two procedures is the stepwise procedure. All three methods rely on the significance levels of the coefficient estimates and linear correlations between the potential independent variables and the dependent variable to make the selection.

As will be seen later, the best model selected from stepwise regression often has a mulicollinearity problem. Nevertheless, the set of variables identified from this process can be a good starting point.

7.1.1.3 Factor Analysis

The purpose of factor analysis is “the orderly simplification of a number of interrelated measures (Child, 1990).” In other words, it helps researchers to determine whether the correlations between a set of observed variables can be explained in terms of a small number of artificial, or latent, variables called "factors". In addition, by examining the "loading", or the linear correlation between each variable and a factor, researchers can identify one or more variables that have a strong relation with each factor. APPENDIX C describes the steps of the factor analysis.

Sometimes, a procedure called factor rotation is carried out to explicitly determine the factors. The rotation reformulates the factors so that the loadings on the few initial variables are as large as possible (Rabe-Hesketh and Everitt, 2003). Rotation is mostly used to create a set of loadings that are more interpretable. In this study, rotation was performed with a constraint that the resulting factors are orthogonal to each other. Finding influential factors and examining the loadings are helpful exercises when constructing a regression model, since the researchers may be able to avoid selecting redundant independent variables.

7.1.1.4 Model Selection Criteria

Throughout the multiple regression analyses, four criteria were used for evaluating the models. First and foremost, as discussed before, multicollinearity should be avoided. Second, the signs of the regression coefficients should be reasonable. Third, high R-squared and adjusted R-squared are desirable. Fourth, additional goodness of fit tests should be used to supplement the R-squared and adjusted R-squared. Long and Freese (2003) at Indiana University have created a STATA command that calculates various post-estimation statistics that include the criteria for evaluating competing models. Among them we used Akaike's Information criterion (AIC) and Bayesian Information Criterion (BIC). Like adjusted R-squared, AIC imposes a penalty to the model when adding an additional independent variable. Specifically, the deviance of the model, which is the log-likelihood of the model multiplied by -2, is adjusted by adding the number of parameters in the model times two. When comparing competing models, the one with the lowest AIC is preferred. The BIC is also used for model selection. Like AIC, it considers the effects of the number of independent variables in the model. BIC is calculated by adding the natural logarithm of the sample size multiplied by the number of parameters in the model. Thus, the smaller the BIC, the better the model fit.

7.1.2 Building Predictive TTG Models

In contrast to the model for delivery units, the TTG models that predict the number of deliveries could not be constructed with the ordinary least square (OLS) technique due to its unique distribution. Since the delivery frequencies are clearly "count" data, the Poisson and Negative Binomial models were initially used. However, as discussed in APPENDIX F, neither model produced good results for the Furniture Chain A dataset because of the unique distribution of the dependent variable with only two possible outcomes, one or two deliveries per week. Both Poisson and Negative Binomial regressions compute the probability of zero delivery, making those models unreasonable for the data set.

For this reason, binary logit model was used. The model is arguably the most well known model for analyzing discrete choice situations. The model assumes that the probability of choosing one alternative over the other is a function of the independent variables. For example, in a two-mode situation, the probability of driving a car over the other mode, say a bus, can be formulated as a function of travel time, income, and other variables that capture the characteristics of the alternatives as well as the traveler demographics. For this research, it is assumed that the number of deliveries that a store receives is determined by the decisions that the Furniture Chain A has made as a part of business operation strategy. While it is not possible to model the operation strategy directly, it is hypothesized that it can be approximated by store-specific and socioeconomic variables.

7.1.2.1 Model Selection and Validation

Several selection criteria were used for the logit model. First, like the linear regression, redundancy among independent variables should be avoided. Second, the signs of the regression coefficients should be reasonable. Third, sensitivity and specificity of the model were taken into consideration. Sensitivity is the probability of correct predictions of positive events (coded as 1), which is two deliveries per week for this research. On the other hand, specificity is the probability of correct predictions of negative events (coded as 0), which is one delivery per week for this research. Thus, the dependent variable for each store in the dataset is coded as either 0 or 1. Sensitivity and specificity measure the ability of the model to replicate both majority and non-majority responses. In addition, the overall probability of correct predictions was considered. Fourth, the receiver operating characteristic (ROC), shown in Figure 7-2, was considered. The ROC curve shows the tradeoff between the probability of correct classification of positive events (sensitivity) and the probability of false positives (i.e. incorrect classification of negative events), which is 1-specificity. The area underneath the graph indicates the predictive power of the model. When the area underneath the line, called the "ROC curve" is 1, it indicates perfect predictive power. The ROC curve is constructed by connecting the coordinates (y-value is sensitivity and x-value is 1-sensitivity) for various values of "cutoff points", or in this case, the threshold value of the probability predicted by the logit model to predict the response. For example, in the upper-right corner of the plot, where the sensitivity is 1, “1-specificity” is also 1. This means that by setting the threshold probability value for predicting the stores with two deliveries per week to zero, all the stores are predicted to receive two weekly deliveries.as. This will result in correct prediction for all the stores that actually receive two deliveries per week, but none of the stores with once a week delivery will be predicted correctly. This graph can also be used to determine the appropriate cut-off value to make binary classifications. For this research, the cut-off value of 0.5 is used. In the binary case, the cut-off value of 0.5 seems to be intuitively appropriated. For the fifth criterion, AIC and BIC methods discussed previously were also used.

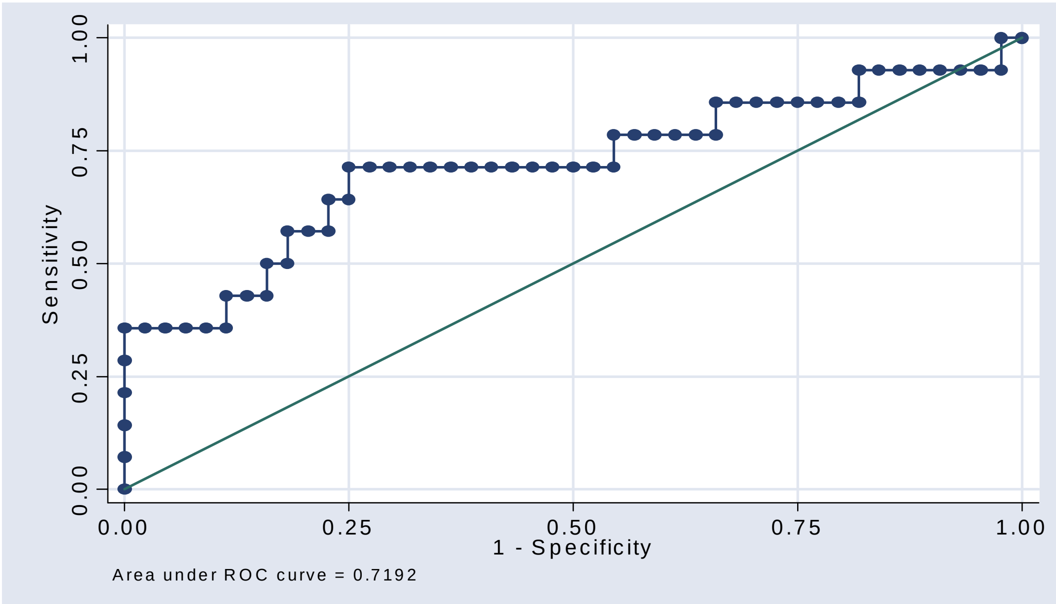


Figure 7-2 Receiver Operating Characteristics Curve

7.2 Model for the Amount of Merchandise Replenished

Two datasets, from Furniture Chain A and Shoe Chain A, were examined separately using the multiple regression. The purpose of the analysis was to identify the relationships between the amount of merchandise delivered and store-specific information and socioeconomic characteristics. The aim of these efforts is to ascertain whether it is possible to estimate the demand for merchandise, as captured by the weekly number of delivery units to each store, from observable variables. Since the demand for merchandise should be intimately related to the number of truck trips made to each store, quantifying demand will greatly facilitate the estimation of the TTG.

As discussed in the previous chapter, the analysis consists of four steps. Before proceeding to the detailed discussion of each model, it should be mentioned again that the standard regression assumptions are not relevant for this dataset since the data points came from a single chain. In other words, two datasets represent population, i.e. all the stores in the territories of the surveyed distribution centers. Thus, standard requirements for a linear regression such as homoscedasticity, normality and others are not meaningful in this case.

7.2.1 Furniture Chain A

7.2.1.1 Models with Store-specific variables

The four best models examining the impacts of store specific variables are exhibited in Table 7-2. These models were developed through lengthy try-and-error process that involved a number of different combinations of transformed and non-transformed independent and dependent variables. Throughout the process, R-squared and the reasonableness of the signs of coefficients were used as the key measures of the effectiveness. Log-transformation of the dependent variable seemed to improve the model's performance, thus all the models in Table 7-2 are fitted for "logwkpallet" (log transformed number of pallets per week) as the dependent variable. The levels of fit are consistently poor. However, some interesting findings were produced. The signs of the coefficients are reasonable except for *sqsale* (squared sales) in Model 2. Model 4, which included only the store type and *logemp*, log transformed *employment*, produced the best adjusted R squared. Model 3 that included store location variable, *mal1l*, and Model 1 that included all the variables, show lower adjusted R-squared in comparison. Thus, adding those variables does not improve model's performance. The number of employees was the most effective variable for explaining the variation in the delivered merchandise. On the other hand, *sqsale,* which had a strong linear relationship with the number of employees, seems mostly irrelevant, a surprising result.

Table 7-2 Regression with Store-Specific Variables - Furniture Chain A

Variable	Model 1	Model 2	Model 3	Model 4
logemp	0.0457	0.0527	0.0561	0.0455
sqsale	0.0006	-0.0004		
mall1	0.0455		0.0440	
stchar1	0.1610	0.1776	0.1602	0.1785
stchar3	0.3023	0.3291	0.3017	0.3302
constant	2.1289	2.1690	2.1313	2.1681
R-squared	0.1406	0.1373	0.1406	0.1373
Adj. R-squared	0.0580	0.0722	0.0757	0.0894

Dependent variable: *Logwkpallet*

7.2.1.2 Full models

The models discussed in this subsection use the socioeconomic variables with four different types of weight (including no weighting) for socioeconomic variables as well as the store-specific variables. The best model was identified for each of the weighting schemes and compared. As mentioned earlier, Factor analysis and Stepwise regression were used to filter the socioeconomic variables.

Table 7-3 compares the best model from each of four weighting schemes. For all the models, the dependent variable is the log of the number of pallets delivered per week to stores (*logwkpallet*). These models were selected based on various criteria including: R-squared values, reasonableness of coefficients, and absence of multicollinearity among independent variables. As far as R-squared and adjusted R-squared are concerned, the first model, unweighted, produced the best results. Also, Akaike's information criteria (AIC), for the first model is the smallest, indicating it as the best fit model. In addition, Bayesian Information Criteria (BIC), of the first model is the smallest. Taking into account all these criteria, the unweighted model, Model 5, seems to be the best model. However, the negative sign of the coefficient for the *logemp* is problematic. It is obviously unintuitive and also inconsistent with all other models including the ones reviewed in the previous section. For this reason, Model 5 was not identified as the best model.

Model 6 performed second best in all criteria except for BIC. The signs of the coefficients are reasonable. This model shows positive effects of the number of employees and median income, as well as the negative impact of low-income population on the number of pallets delivered. The interpretation of median age is difficult since it is transformed (as an inverse of the square root of the median age). The positive coefficient estimate for this variable implies a negative association with the dependent variable. It should be noted; however, that this does not directly mean a higher median age of the market shed is related to a lower amount of merchandise delivered because this variable is weighted by the inverse of the distance between the store and each Census block group in the market shed. Thus, the influence of the spatial distribution of the elderly (or young) population within the market shed, might have produced the positive coefficient. Finally, location characteristics and store types are related to the weekly volume of pallets delivered to the stores. These results agree with the findings from the preliminary interviews documented in Chapter 6.

Model 7 and 8 also produced reasonable signs of coefficients. However, except for BIC, these models are inferior to the Models 5 and 6. As a result, the Model 6 was identified as the best among four alternatives.

Table 7-3 Comparing Four Best models - Furniture Chain A

	Model 5	**Model 6**	**Model 7**	**Model 8**
	Weights			
Variables	*No weight*	$\frac{1}{Dist}$	$\frac{1}{Dist^2}$	$\frac{1}{Dist^3}$
logemp	-0.0437	0.0427	0.0023	0.0032
sqrt(density)	-0.0012			
(median age) cubed	6.66E-06			
1/(average househod size cubed)	4.3621			
sqrt(median income)	0.0042			
mall1	0.0344	0.0555	0.0156	0.0118
stchar1	0.2450	0.1693	0.1897	0.1914
stchar3	0.3642	0.1626	0.1563	0.1482
1/sqrt(median age)		36.8899		
median income		2.20E-09		
low income population		-2.30E-09		
log(african american)			-0.0393	-0.0438
log(average household size)			0.0113	
log(median income)				0.0294
constant	0.7740	1.9014	3.0343	2.6861
R-squared	0.3223	0.2616	0.2110	0.2179
Adj R-squared	0.2117	0.1583	0.1182	0.1259
AIC	0.510	0.561	0.593	0.584
BIC	-187.3	-186.5	-186.7	-187.2

7.2.1.3 Findings

- The model using the inverse of distance as the weight for socioeconomic variables was identified as the best overall performer.
- When considering only the store-specific variables, it was found that the log of *employee*, store locations and types are positively associated with the log of the number of pallets delivered to stores. The coefficients indicate that off-mall store location, and also the combo and outlet stores tend to receive higher number of pallets per week. Meanwhile,

the model suggests that it is unlikely that *sales volumes* are related to the number of pallets delivered.

- Log transformation of the dependent variable, weekly pallets, produced better results regardless of the weights used.
- Store specific variables, *employee, sales, size, mall,* and *stchar,* alone cannot sufficiently explain the number of pallets delivered. However, adding these variables to the models with socioeconomic variables produced better results in terms of R-squared.
- Lastly, as the DC manager identified, the number of pallets delivered is positively associated with the median income and negatively associated with the number of low income people in the market shed.

7.2.2 Shoe Chain A

In this section, the development of the models for Shoe Chain A is discussed. Shoe Chain A has one distribution center that covers 259 stores in 23 states. The dependent variable was the number of cartons of shoes delivered to each store per week. Only two store specific variables, employee and (*employee*) sales volume (*sales*), were available. However, like Furniture Chain A, 37 socioeconomic variables were considered. After removing missing values, 207 stores were used for the regression analyses. As was for Furniture Chain A, four types of weights were applied to socioeconomic variables. As was in previous sections, the first series of regressions used only the store-specific variables. Then, the full models were developed.

7.2.2.1 Models with Store-Specific Variables

Table 7-4 presents three possible models using two store-specific variables. Transformation was not performed since the dataset fitted well with linear models. The levels of fit, judged by R-squared, are considerably higher than that for the Furniture Chain A models. However, Model 9 shows negative coefficient estimate for *Sales* variable. This is partly caused by the fact the there was a strong linear correlation between *employee* and *sales* as shown in the plot included in APPENDIX D. The comparison between Model 10 and Model 11 indicates that *employee* is a strong determinant of the number of cartons delivered. Meanwhile, *sales* seem to have a very weak association with the number of cartons delivered.

Table 7-4 Regression with Store-Specific Variables - Shoe Chain A

	Model 9	Model 10	Model 11
employee	6.0101	5.8551	
sales	-0.0011		0.0103
Constant	206.5956	206.2482	263.9847
R-squared	0.3942	0.3935	0.0897
Adj R-squared	0.3883	0.3905	0.0853

7.2.2.2 Full Models

As described earlier, the building of the full model, with both store-specific and socioeconomic variables, was performed using Factor Analysis and stepwise regression as the starting points. However, those two techniques did not provide useful directions in the selection of the independent variables since stepwise regression resulted in no independent variable (i.e. all the variables were rejected), and Factor Analysis failed to identify strong factors. In general, there were considerable overlaps among the factors. Due to these difficulties, the development of the models relied on the trial and errors process to find the best models for all four weights.

Table 7-5 summarizes the best model under each of four weighting schemes. Dependent variable for all the models was the number of cartons delivered to each store per week. All the models performed considerably better than the models for Furniture Chain A. However, it should be noted that the model fit did not improve significantly from the addition of socioeconomic variables. In fact, R-squared of the model that include only *employee* (0.391) is not much different from those shown in this table, underscoring the importance of the number of employees as the explanatory variable for the amount of shoes sold.

Although the R-squared value of Model 12 is the highest, the differences among the models are marginal. However, Model 12 seems to be the best performer since it shows the best scores in all the criteria. The sign of *median age* seems reasonable, since the chain focuses on the younger people as potential customers. The positive coefficient for the variable, *youth*, also supports this interpretation.

Table 7-5 Comparing Four Best Models - Shoe Chain A

Variables	Model 12	Model 13	Model 14	Model 15
employee	5.5516	5.7130	5.7939	5.8009
high income		0.0000		
median age	-2.3948	0.0003		
medina income		0.0000	0.0000	
density			0.0000	0.0000
white			0.0000	0.0000
black			0.0000	0.0000
1/(avg. hh. Size) cubed	757.0660			
youth	0.0002			
Constant	127.9035	211.0439		124.2188
R-squared	0.4247	0.4090	0.4042	0.4050
Adj R-squared	0.4133	0.3973	0.3893	0.3933
Mean VIF	1.1800	5.0000	2.8200	1.2100

7.2.2.3 Findings

- The unweighted model performed best. This is mostly a result of the weak influences of socioeconomic variables, to which the weights were applied.
- The number of employees is the key independent variable that explains the variation in the number of shoe cartons delivered to each store. In fact, R-squared of the model that includes only *employee* is not much different from that of the full models.
- The age of the population within the market shed seems to have a negative impact on the number of shoes delivered. This is not surprising since Shoe Store A targets younger population as the market.

7.2.3 Summary of Store-Specific Variable Models

Regression analyses for Furniture Chain A and Shoe Chain A were carried out. In terms of R-squared, the model for Shoe Chain A, performed significantly better, indicating stronger linear relationships between *cartons* and selected independent variables. Also, both models suggest the number of employees is the key independent variable that greatly affects the fit of the model. This is more so for Shoe Chain A, than for Furniture Store A. In contrast, we could not find clear evidence of the relationship between *sales* and the amount of merchandise delivered to the stores. Finally, probably the most significant finding is that for Furniture Chain A, the variables related to store type and location did not have as strong impact as we expected despite the fact that our data source indicated that those factors were key determinants of the sales.

7.3 TTG Models

This section attempts to build a TTG model using the Furniture Chain A dataset. The model assumes the frequency of delivery trips to the stores (from the DC) is dependent on both the store-specific variables, that are assumed to be the proxies for firm's decision-making behavior, and the socioeconomic characteristics of the market shed, which represent the potential demand for the goods being sold at the stores.

Since the data are for the individual retail stores, each delivery generates two trip ends, inbound and outbound[5]. It should be noted that it is unlikely that the stores included in the dataset receive additional truck trips since all the merchandise replenishment is performed through DCs. The only exception may be the use of parcel carriers to replenish small items, which was identified in the survey. Parcel delivery, is becoming a rather common practice. Our dataset did not include parcel delivery data,

It may seem that if one can predict the amount of merchandise delivered to each store, as we did in the last section, the number of truck trips required to deliver those items can be derived easily. However, that would be the standard commodity-based model. The hypotheses described in Chapter 3 that were validated by the interview with the DC manager contend that the operational decisions by the business determine the frequency of delivery trips even when the demand for the merchandise is fixed.

As opposed to the exploratory nature of the regression analysis described in the previous section, this effort focuses on developing the model with the best predictive capability. In this type of application, once a model is calibrated with the baseline dataset and satisfies the predefined model selection criteria, it should be applied against a separate set of data to test the transferability and predictive capabilities. We used the data from Furniture Chain E for this purpose.

---

[5] Since truck trips can be considered as non-home based, this can be also defined as one attraction and one production trip ends.

7.3.1 Poisson Regression and Negative Binomial Regression

Prior to the construction of the binomial logit models, two other types of models(Poisson and Negative Binomial models), often used to regress the frequency data,, were tested. Even though the dependent variable is certainly a count data, these models were not successful. This is because these models always include zero trips that does not exist in Furniture Chain A's dataset. The Negative binomial model is often used to address the overdispersion problem that is quite common in Poisson models since it assumes that the mean and variance are the same. However, overdispersion was not observed in this dataset, and also the fit of the negative binomial model was not satisfactory. A detailed discussion on Poisson and negative binomial regression analysis with the Furniture Chain A dataset can be found in APPENDIX F.

7.3.2 TTG Model with Store-Specific Variables

This section attempts to build a binary logit model with only the store-specific variables. For Furniture Chain A, they are: the number of employees, annual sales volume, store location characteristics and store types.

Table 7-6 compares five different models. The first three models, Models 16, 17 and 18, that are shown in the three columns on the left, examine the influence of *employee* and *sales* variables. The performance is generally poor. Pseudo R-squared values are low, and more importantly, all three models predict almost all the stores to be one-delivery-per-week type. Thus, the sensitivities[6] (percent of correct predictions for stores with two deliveries per week) are extremely low. Thus, even though specificities (percent of correct predictions for stores with one delivery per week) are it is achieved by sacrificing the sensitivity. As noted previously, estimated probability of 0.5 was used as the threshold value for classifying each store into either one delivery or two deliveries per week. Lastly, receiver operating characteristic (ROC) scores, show adequate, but not excellent, predictive powers of the models. For Model 19, the dummy variables for store location and types were added to Model 16. The dummy variable, *mall1*, indicates the

[6] See Section 7.1.2.1 for the definition of specificity and sensitivity

stores in off-mall locations. Dummy variables, *Stchar2* and *stchar3*, represent conventional type and outlet type stores, respectively.

As shown, in all aspects, Model 19 showed a substantial improvement over the previous three models. It has a much higher pseudo R-squared value. Sensitivity indicates that 50 percent of two-delivery-per-week stores were correctly predicted a remarkable improvement over the aforementioned models. The overall probability of correct classification is almost 88 percent. In addition, the ROC is close to 0.9, indicating a high predictive power. As the performance of Model 20, with only the store type and location variables, indicates, these variables are effective predictors of delivery frequencies. Even though the pseudo R-squared is lower than that of Model 19, it predicts the delivery frequencies almost equally well. In fact, the model's performance show that with the knowledge of store type and location alone, one can correctly predict the delivery frequency for over 85 percent of the stores.

Table 7-6 TTG Model with Store-Specific Variables - Furniture Chain A

Variables	Model 16	Model 17	Model 18	Model 19	Model 20
Employee	0.0352	0.0805		-10.52388	
Sales	3.47E-07		5.06E-07	5.38E-05	
mall1				0.6583	0.2406
stchar2				-109.0505	-3.8885
stchar3				-107.6011	-3.6354
Constant	-2.0525	-1.8837	-1.9961	104.6751	2.0086
Pseudo R2	0.0211	0.0137	0.0199	0.4536	0.2853
Sensitivity	7.14%	0.00%	7.14%	50.00%	50.00%
	(1/14)	(0/14)	(1/14)	(7/14)	(9/18)
Specificity	100.00%	97.73%	100.00%	100.00%	98.04%
	(44/44)	(43/44)	(44/44)	(44/44)	(50/51)
% correct	77.59%	74.14%	77.59%	87.93%	85.51%
classification	(45/58)	(43/58)	(45/58)	(51/58)	(59/69)
ROC	0.5909	0.5950	0.5698	0.8969	0.7696
n = number of obs.	58	58	58	58	69

Compared to the first three models, all aspects of the predictive power are superior for the last two models. In addition, the signs of the coefficient estimates indicate that the stores in off-mall locations are more likely to receive two deliveries per week. Also, compared to the combo stores (*stchar1*), conventional (*stchar2*) and outlet (*stchar3*) stores are less likely to receive two deliveries per week. This implies that information on location and store types improves the prediction of the frequency of truck trips. The data shown in the cross-tabulation, Table 7-7 support this finding. The number in each cell represents the share among the stores with the characteristics indicated by the column and row headings that receive two deliveries per week. For example, the figure in the "combo/off-mall" cell indicates that 87.5 percent of the combo stores located in off-mall locations received two weekly deliveries. The row totals (the last column) show that 34 percent of stores in off-mall locations received two deliveries per week, whereas it is only 17 percent for the stores in malls. Also, the column totals (the last row) show that 90 percent of combo stores, 14 percent of conventional stores and 20 percent of outlet stores received two deliveries per week. The data indicate that combo stores are highly likely to be two-delivery-per-week type, compared to other two store types. Furthermore, all combo stores in a mall receive two delivery trips without exception.

Table 7-7 Share of Stores Receiving Two Deliveries per Week

		Store Types			Total
		Combo	Conventional	Outlet	
Location Characterist	Off-mall	0.875	0.176	0.200	0.343
	Mall	1.000	0.125	0.000	0.176
Total		0.900	0.143	0.200	0.261

7.3.3 TTG Model

Like the multiple regression analysis, four types of weights for socioeconomic variables were tested. However, the models using two of the weights, inverse of distance squared and inverse of distance cubed, performed extremely poorly, and were eliminated. Since these weights severely penalize the Census block groups that are far from the store, this poor performance suggests that the socioeconomic characteristics of the immediate vicinity of the store do not play an important role in determining the number of delivery trips.

The selection of the independent variables for the logit model relied on the results from the two variable selection techniques, Factor Analysis and Stepwise regression that were performed for the multiple regression analysis. Multicollinearity diagnostics were used to remove the variable with the highest VIF. The predictive powers of the models were used as the primary guidance for selecting the best model.

After examining a number of potential combinations of independent variables, the two models, shown in Table 7-8 were found to be the best models. In terms of pseudo R-squared, Model 21 is superior; however, since the purpose of this effort is to derive the model with the best predictive capability, much emphasis should be placed on the overall percentage of correct predictions. Model 22 correctly estimated about 91 percent of observations, while Model 21 successfully classified about 88 percent. In addition, sensitivity and specificity figures for Model 22 are slightly better. On the other hand, the ROC is better for Model 21.

The sensitivity values for the models indicate weakness in correctly identifying the two-deliveries-per-week stores. These models were able to identify only 55% to 65% of those stores. Thus, if these models are used in practice, the TTG will likely to be underestimated.

Table 7-8 Final TTG Models - Furniture Chain A

	Model 21		Model 22
Weight	no weight	Weight	Inverse of distance
N	58	N	58
Variables	Estimates	Variables	Estimates
		sales	1.18E-06
employee	0.1242	mall1	0.0858
mall1	0.9558	stchar2	-22.0636
stchar2	-22.4117	stchar3	-463
stchar3	-22.4725	w1pop00_sqmi	8.62E-07
highincome	0.0001	w1lowincome	-6.47E-08
med_age	0.4602	w1med_age	-0.0000
sqmincome	0.0134	w1aincome	-1.00E-15
Constant	-3.7462	Constant	18.1270
Pseudo R2	0.5602	Pseudo R2	0.5076
Sensitivity	57.14%	Sensitivity	64.29%
	(8/14)		(9/14)
Specificity	97.73%	Specificity	100%
	(43/44)		(44/44)
% Correct	87.93%	% Correct	91.38%
Classification	(51/58)	Classification	(53/58)
ROC	0.9221	ROC	0.9042

It also should be noted that these two models show only marginal improvements over the store-specific variable models that were presented in Table 7-6. The performance of the best model from that table is almost identical to that of Model 21, which includes socioeconomic variables. This underscores the importance of store-specific variables and also suggests that socioeconomic variables of market shed are not critical for predicting the truck delivery frequencies. The signs of coefficient estimates are reasonable. They indicate positive impacts of: income level, population density, the number of employees, and age. The coefficients of the store type dummy variables point to the more frequent replenishment for Combo type stores. Also, off-mall stores, coded by the *mall1* dummy, tend to receive more frequent deliveries.

Table 7-9 shows the performance of several variations of Models 21 and 22. These models were constructed to assess the importance of store-specific variables. For example by comparing Model 21 with its variant without the *employee* variable, it is possible to determine its contribution to the predictive power of the TTG model. The analyses show that the number of employees and sales volumes do not play a critical role in predicting the delivery frequency by trucks.

Table 7-9 Variants of TTG Models

	Model 21	W/O employee	W/O types		Model 22	W/O sales	W/O types
Weight		No weight		Weight		Inverse of distance	
N	58	69	58	N	58	69	58
Variables		Estimates		Variables		Estimates	
				sales	1.18E-06		1.02E-06
employee	0.1242		0.1003	mall1	0.0858	-0.1501	1.034
mall1	0.9558	1.0281	1.1691	stchar2	-22.0636	-4.2704	
stchar2	-22.4117	-5.7744		stchar3	-463	-3.9247	
stchar3	-22.4725	-4.1760		w1pop00_sqmi	8.62E-07	6.48E-07	8.29E-07
highincome	0.0001	0.0000	0.0001	w1lowincome	-6.47E-08	-8.04E-08	-5.79E-08
med_age	0.4602	0.6261	0.2423	w1med_age	-0.0000	-0.0000	-0.0001
sqmincome	0.0134	0.0212	0.0040	w1aincome	-1.00E-15	2.20E-13	9.72E-14
Constant	-3.7462	-27.4889	-14.6557	Constant	18.1270	2.4357	-3.0178
Pseudo R2	0.5602	0.4893	0.2131	Pseudo R2	0.5076	0.4079	0.2182
Sensitivity	57.14%	61.11%	35.71%	Sensitivity	64.29%	66.67%	21.43%
	(8/14)	(11/18)	(5/14)		(9/14)	(12/18)	(3/14)
Specificity	97.73%	96.08%	93.18%	Specificity	100%	98%	95%
	(43/44)	(49/51)	(41/44)		(44/44)	(50/51)	(42/44)
% Correct Classification	87.93%	86.96%	79.31%	% Correct Classification	91.38%	89.86%	77.59%
	(51/58)	(60/69)	(46/58)		(53/58)	(62/69)	(45/58)
ROC	0.9221	0.9096	0.7825	ROC	0.9042	0.8355	0.8052

Removing those variables from the TTG models reduced pseudo R-squared and ROC, the accuracy of the prediction, measured by the numbers in the bottom four rows in the table, decreased only slightly, or in some cases, improved. In fact, the sensitivities of Models 21 and 22 actually improved slightly when *employee* and *sales* were removed, respectively. It should be noted that the exclusion of these two variables enabled us to use larger datasets, since some of the data points were missing employee counts and sales information. Therefore, a direct comparison of the performance measures requires cautious interpretation.

In comparison, store type seems to play a crucial role in the performance of the TTG models. The pseudo R-squared drops considerably when the dummy variables for store types are removed from the models. Predictive power, especially the correct classification of two-deliveries-per-week stores, also decreases by a wide margin.

The data for Furniture Chain E stores in the Chicago area were used for validation and also for testing the transferability of the models. As mentioned in Chapters 4 and 5, of the 11 stores in the Furniture Chain E dataset, some stores receive one delivery per week; and others receive two deliveries. Thee store-specific variables were available, *stchar1* (off-mall) *employee* and *sales.* However, all stores are located in off-mall locations. Also, since all the stores sell nearly identical mixes of merchandise, store types could not be used as an independent variable. Thus, it was not possible to conduct this test for Models 21 and 22, as both include store types as independent variables. Instead, two of the variants that are shown as "W/O types" in Table 7-9, and also two of the best the models that do not include store type or location were tested using the Furniture Chain E dataset.

The results, shown in Table 7-10 indicate that while all four models performed at approximately the same level for the Furniture Chain A dataset, differences became apparent when they were applied to Furniture Chain E. The models that include weighted socioeconomic variables outperformed un-weighted ones. The former classified approximately 64 percent of the delivery frequencies for Furniture Chain E stores while the latter succeeded only about 30 percent of time. Another notable result is that, in contrast to the cases for Furniture Chain A, the models generally overestimated the two-deliveries-per-week stores. Thus, while the models produced decent sensitivity values, their ability to correctly identify one-delivery-per week stores was disappointing. This phenomenon is most likely related to the fact that all of Furniture Chain E's stores in the dataset are off-mall types. One plausible interpretation is that while the location type was one of the factors that were used by Furniture Chain A to determine the replenishment frequency, this was not the case for Furniture Chain E. Furniture Chain A may be using other criteria that are not captured by the models presented. Thus, evidence indicates that the transferability of the models from Furniture Chain A to Furniture Chain E must be questioned even though they are close competitors that sell similar merchandise.

Table 7-10 Validation and Transferability Analysis

		Model 23	Model 24	Model 21 W/O types	Model 22 W/O types
	Independent Variables	employee, highincome, med_age, sqmincome	sales, w1pop00_sqmi, w1lowincome, w1med_age, w1aincome	employee, highincome, med_age, sqmincome, mall1	sales, w1pop00_sqmi, w1lowincome, w1med_age, w1aincome, mall1
	Weight	no weight	Inverse of distance	no weight	Inverse of distance
Furniture Chain A	Sensitivity	35.71% (5/14)	21.43% (3/14)	35.71% (5/14)	21.43% (3/14)
	Specificity	100% (44/44)	97.73% (43/44)	93.18% (41/44)	95.45% (42/44)
	% Correct classification	84.48% (49/48)	79.31% (46/58)	79.31% (46/58)	77.59% (45/58)
Furniture Chain E	Sensitivity	33.33% (2/6)	83.00% (5/6)	66.67% (4/6)	100.00% (6/6)
	Specificity	20.00% (1/5)	40.00% (2/5)	0.00% (0/5)	20.00% (1/5)
	% Correct classification	27.27% (3/11)	64.00% (7/11)	36.36% (4/11)	63.64% (7/11)

7.4 Summary of Findings

Since the dependent variable for the Furniture Chain A dataset had only two outcomes, the binary logit model was the most appropriate for this study. Neither Poisson nor negative binomial regressions produced satisfactory levels of fit. However, this finding may not be applicable to other datasets.

Store location and store types are the most influential variables for TTG models. Inclusion of those variables always improved model's predictive power. Of all the variables examined in this study, the store type was identified as the most important predictor of the delivery frequency. Although the models that include the store type variable could not be validated using the Furniture Chain E dataset, it is reasonable to expect that such model would outperform four models that were tested. It should be noted that location and store types are characteristics that are easily observable while many of the variables that were examined in this study can not be obtained easily.

The number of employees, which is often used as the predictor of trip generation, showed a positive association with the delivery frequency. However, it has only a marginal effect on the predictive power of the TTG models.

Socioeconomic characteristics of market shed, while providing some improvement in the performance of the TTG models, are far less important than the store type and location.

8 . SUMMARY AND IMPLICATIONS OF STUDY

8.1 Summary of the Study

The goal of this study was to develop a new generation of truck trip generation (TTG) model that takes into account the supply chain management and logistics strategies of businesses. As the initial task, to facilitate the development of the data collection strategy and also the conceptual framework of the TTG model, experts from the freight and logistics industries were interviewed. The interviews directed our research toward retail sectors that sell relatively homogeneous merchandise in terms of weight-to-volume ratio, inventory turnover and supply chain system. This eliminated some sectors such as grocery stores.

Data were collected using a combination of instrument-based survey, store visits, and phone calls. While the data collection effort encountered a lack of cooperation from the businesses, in the end, we obtained datasets from two national retail chains (a shoe retailer and a furniture chain), containing a total of 335 stores, that included detailed information on the supply chain strategies and some store characteristics. In addition, supplemental data were obtained from a total of 94 stores belonging to seven different national retail chains.

Using the obtained data, three clusters of analysis were conducted. The first cluster used graphical tools and descriptive statistics to examine the broad trends in the datasets. This analysis identified strong association between store-specific characteristics such as store and site types and the frequency of truck trips to each store for delivering replenishment merchandise.

Statistical analysis of the data involved multiple regression and logit model. The former was used for investigating the relationship between the characteristics of the store itself, the market shed, and the amount of merchandise delivered to each store every week. The latter was used to develop the models that predict the number of truck trips between the retailer's distribution center and stores. It should be noted that the data collected in this study did not allow the use of standard inferential statistics, such as hypothesis tests based on t-statistics. This is because the data, while having many data points, came only from two retailers, and did not constitute a sample. Each dataset included all the stores that were replenished by the distribution center from which the data were obtained. Thus, the goodness of fit measures such as adjusted R-squared, ROC, and proportion of accurate prediction of the dependent variable were used as the main model development and selection criteria.

It was found that the number of employees is an important factor in the estimation of the amount of merchandise delivered to each store. However, the role of the sales revenue was not clear. For the furniture retailer, Furniture Chain A, the role of the store and location types on the amount of merchandise delivered was not strong, contradicting the comments made by a company employee.

A logit model was applied only to the Furniture Chain A's data set due to data limitations. The best model was able to correctly identify over 90 percent of the data points in the calibration dataset (for Furniture Chain A). However, the model performed only adequately, at 64 percent correct, in identifying the stores that received two deliveries per week. The model that used weighted socioeconomic variables of the market shed performed slightly better than the model that used un-weighted socioeconomic variables. When a variant of the best model[7] was applied to the data from a competing furniture chain (Furniture Chain E), it correctly identified 64 percent of the delivery frequencies. The inconsistency of the results, and also the fact that Furniture Chain A and E are very similar in terms of their merchandise and target population suggest that a guideline or classification taxonomy to determine the transferability of the models are needed.

The most notable finding was that including the variables for store types and location always improved the models' performance considerably, while sales and employee count only had marginal effect. In fact, the logit model that included only the store location and type predicted 85.5 percent of the delivery frequencies correctly for the calibration dataset.

8.2 Implications of the Study

The knowledge gained during the data collection process brought valuable insights into the relationship between the business strategies concerning logistics and supply chain management and the TTG characteristics of businesses. The knowledge building process also taught us that a closer cooperation between the private sector and the academia is absolutely critical for the development of better analytical tools to address increasing truck trips in urban areas.

---

[7] Store type had to be excluded from the model since the Furniture E dataset did not include such variable.

The findings from the model building generally supports the conceptual TTG model for big-box retail chains that was presented in Chapter 3 except for the role played by the customer demand. We found that independent variables such as the number of employees and floor space, which have been used as the proxies for store volumes are not necessary good predictors of truck trips. This finding suggests that the current methods for estimating the truck trips generated by newly proposed retail stores, e.g. ITE trip generation rates, are prone to severe errors.

Table 8-1 compares numbers of trip ends estimated by our models to actual number of trip ends and also the figure calculated based on the ITE Trip Generation (ITE, 2003). The detailed description of the calculation using the ITE method is given in APPENDIX G.

Table 8-1 Comparison of TTG Models Versus ITE Trip Generation Method

	Furniture Chain A	Furniture Chain E
Number of stores	58	11
	Total number of truck trip ends per week	
Model 23 (I-1)	126	34
Model 24 (II-1)	124	38
Model 21 W/O types (I-4)	132	40
Model 22 W/O types (II-4)	126	42
ITE (based on number of employees)	2265	588
Actual	144	34

The comparison reveals that the ITE method severely overestimated the truck trip ends while our TTG models produced reasonably accurate estimates. In particular, our models performed surprisingly well for Furniture Chain E considering the fact that, as discussed in the previous chapter, a relatively small portion of the stores was correctly categorized into one or two-delivery per week. This is due to the fact that even though the models made errors in categorizing the stores, the errors often canceled out each other, and in the end, the total number of stores in each category was close to the actual data.

While the errors associated with the ITE method seem preposterous, in retrospect, they are not too surprising. Firstly, there is no clear definition of "trucks" given by the ITE manual. As discussed in the ITE's "Recommended Practice" (ITE, 2003b), "[t]here are, figuratively speaking, as many definitions of the term "truck" as there are potential uses of truck trip generation data." Thus, the "five percent" figure given by the ITE manual as the share of trip ends associated with trucks may include light duty trucks and parcel delivery vans, which was not included in our data. Secondly, the data in the ITE manual contained only 8 data points. The data also contain both traditional stores and warehouses with showrooms. It is difficult to develop accurate and transferable trip generation model based on such a small dataset with a wide variation among the data points using the linear regression technique. Thirdly, the average number of employees per store was 33 in the ITE data. On the other hand, the average number of employees for our data set was 9.24. Even the largest store in our dataset employed only 24 people. The difference can be explained by the fact that the ITE data were collected in "the late 1970's through mid-1980's" (ITE, 2003), when the business of furniture retail, including the supply chain system, was quite different.

Today, most retail chains use large trucks such as 80,000 pound tractor-trailer units to replenish the stores. They also employ highly standardized routing schedules. Under such system, as long as the volume of merchandise sold at a store is less than the capacity of the trucks, which was always the case in our datasets, businesses are able to set the frequency of replenishments independently from the volume of sales, which may be correlated to the variables used by the ITE method (e.g. number of employees or floor areas). This is especially true today because delivery trucks travel hundreds or even thousands of miles to cover stores in different cities, and routing plans are designed to maximize the efficiency for the entire delivery network. Under such conditions, it is generally not economical to make additional deliveries to a small number of stores. For example, it was found that Shoe Chain A never alters the frequency of shipment for any of the stores even during the peak season. The volume of delivery, not the frequency, was varied to meet fluctuations in demand. Also, of the nine retails chains examined, only two used "discriminating" delivery schedule (i.e. more frequent delivery for some stores). Other chains used fixed number of replenishment for all the stores.

In contrast, the variables concerning the routing schedule decisions were found to play a key role in the prediction of delivery frequencies. During the data collection process, we found that Furniture Chain A's strategy for the routing schedule is partly dependent on the types of stores and also location. The company’s overall business strategy is to focus on particular types of customers and products. Placing a greater strategic emphasis on the stores can help accomplish this strategy. Thus, the core assumption that facilitated the development of the TTG models discussed in this report, the presence of a close relationship between the truck trip frequency and business decisions, has been verified.

It should be noted that the data on store and location types are much easier to obtain from observation than the number of employees or store size. Since different retailers may utilize different rules and logic to strategically determine the replenishment frequency, there is no guarantee that those variables are applicable to other cases. However, at a minimum, we have identified a need to identify the variables that capture business strategies for logistics and supply chain management.

Due to data limitations, this study was not able to cover a wide range of retailers. Our analysis found that a model that was developed for a particular retailer based on a strategic decision making process may not be effective for another retailer even if they are close competitors. We also found that classifications such as SIC, NAICS, or those used by private firms did not offer effective taxonomy of the businesses in terms of their TTG characteristics. Thus, for advancing the new-generation TTG model to the application stage, the development of appropriate classification system is imperative. Since such effort will require an enormous amount of detailed data covering many sectors, it may not be achievable without strong support from the public sector and industry trade organizations or individual companies.

Appendix

APPENDIX A. CLASSIFICATION OF RETAIL SECTORS

Apparel & Accessories Retail	**Home Furnishings & Housewares Retail**
Footwear & Related Products Retail	**Home Improvement & Hardware Retail**
Auto Parts Retail	**Jewelry & Watch Retail**
Automobile Dealers	**Military & Government Exchange Retail**
Building Materials Retail & Distribution	**Music, Video, Book & Entertainment Retail**
Camera & Optical Goods Retail	**Musical Equipment Retail**
Computer & Software Retail	**Nonstore Retail**
Consumer Electronics & Appliances Retail	Catalog, Mail Order & Television Sales
Convenience Stores & Truck Stops	Direct Selling
Cosmetics, Beauty Supply & Perfume Retail	Internet Retail
Department Stores	**Office Products Retail & Distribution**
Discount & Variety Retail	**Party & Holiday Accessories Retail**
Warehouse Clubs & Superstores	**Recreational Vehicle, Motorcycles & Boat Retail**
Drug Stores & Pharmacies	**Sporting & Recreational Equipment Retail**
Floor & Window Coverings Retail	Golf Equipment Retail
Floral & Gifts Retail	**Tobacco Retail**
Gasoline Retailers	**Toys & Games Retail**
Grocery Retail	
Natural & Specialty Foods Retail	
Hobby & Craft Retail	

Source: www.hoovers.com

APPENDIX B. SURVEY INSTRUMENT

Dear

As we discussed over a phone call weeks ago, we are collecting data to conduct a research study that is funded by the U.S. Department of Transportation and the Midwest Regional University Transportation Center at the University of Wisconsin, Madison. The goal of this research is to improve the methodology for estimating the number of truck trips coming in and out of business facilities. The formulas currently used by the municipalities and regional agencies to conduct planning studies of proposed industrial and commercial developments are based on outdated data and methodology. The formulas do not reflect the tremendous improvements in the supply chain management that occurred over the last decade. As a result, the traffic impacts of the truck trips associated with proposed developments are inaccurately estimated.

Aside from the contribution to the academic research, we will disseminate the results of this study through various professional channels to help improve the state-of-the-practice. For a more tangible outcome to your business, you will receive the result of the study that includes selected data items that are aggregated to protect the confidentiality. The research team is planning to collect same information from the various types of business; therefore, you will be able to use such information as a benchmark.

Please read the attached data wish list that we prepared to document the types of information we are looking for. We would like to call you in a few days to see if we can expect your cooperation. We are aware that some of the data may be proprietary, but since we cannot make a judgment on our own, we have generated this list without regard to such concern for now. If you can provide even a part of the data in the list, we will be truly grateful.

For the delivery of the information, we will make the arrangement that will work best for you. If you wish, we can provide an UPS account number that you can use to mail the material, or we can visit your facility to pick up the data or make copies of the documents.

All answers and discussions are confidential and will be used only for the purposes of our research. To assure anonymity, if the responses are written in any format, your (personal and business) identity will be coded with an identification number instead of your true name. In addition, the information collected from the survey is not disclosed to others except in an aggregated form.

We would like to thank you in advance for your time.

Kazuya Kawamura	Hyeon-Shic Shin
Assistant Professor	Ph. D. Student
Urban Transportation Center	Urban Transportation Center
University of Illinois at Chicago	University of Illinois at Chicago
412 S. Peoria St. Suite 340	412 S. Peoria St. Suite 340
Chicago, IL 60607	Chicago, IL 60607

BUSINESS AND SITE SPECIFIC TRUCK TRIP GENERATION

DATA WISH LIST

This list includes the data that the research team would like to obtain. As a first step, we are contacting the distribution centers (DCs) because in many cases, the DCs maintain the detailed shipment information to and from each retail store that we are looking for.

I. Truck Trip Information

I-1 Coverage

Information regarding:

(1) The geographical territory that your DC covers

(2) The number of stores attached to your DC function

I-2 Route schedule

Information regarding:

(1) The factors that determine the route schedules of the shipments to the stores

(2) All the route schedule patterns (time and location of each stop, travel distance between each stop, frequency of route per week) used by your DC to replenish the retail stores in Illinois, Indiana, Ohio, and Wisconsin. The information for a longer time period (e.g. one year) is preferred. Otherwise, the data from typical days, weeks, months or quarter will be desired (the questions regarding the seasonal variation in the route schedule is addressed below). We assume that such information will capture the characteristics of truck trip generation in your place and retail stores that you serve.

(3) All the seasonal/special route schedules (time and location of each stop, travel distance between each stop, frequency of route per week) used by your DC, and the time periods you use such schedules.

(4) The factors that determine the replenishment frequency for each store

I-3 Store replenishment

For each of the stores in Illinois, Indiana, Ohio, and Wisconsin that are covered by your DC, we would like to obtain the following data.

(1) The number of replenishment deliveries from your DC per week

(2) The average cubic feet of shipment per delivery

(3) The average weight of shipment per delivery

(4) The average value of shipment per delivery

(5) The sizes of containers used for shipment

(6) The average dwell time per stop

(7) The percentage of shipments delivered by outside vendors or other DCs of your company by trips, cubic feet, weight, value

(8) Schedule for the deliveries by outside venders

(9) The process of the shipments by outside vendors.

II. Retail Store Information

For your DC and each of the retail stores within Illinois, Indiana, Ohio, and Wisconsin

(1) Address

(2) Gross Floor Space (GFS)

(3) Number of employees by month

(4) Number of docs

(5) Truck parking space size

(6) Drop lot space size

(7) Hours of business and operation

III. Discussion: Logistics and Supply Chain Information

It is assumed that the choice of different strategy will influence truck trip generation

(1) What are the factors that determine departure time of shipment?

(2) What types and size of trucks do you have?

(3) In case of for-hire trucks, is there any decision rule for selecting shipment mode between truck-load (TL) and less-than-truck (LTL)?

(4) Many discussions on the new technology and strategies have been made in the areas of supply chain management, logistics, distribution and inventory management. What is the strategy for the efficiency of you distribution network? Do you use any types of DC management tools or technology?

(5) In a supply chain network, the transportation is obviously one of cost components to be minimized. What kind of strategy is used to minimize truck traffic generated in you facility?

(6) Do you use any type of e-commerce such as telephone order and internet? How does it affect your business?

APPENDIX C. FACTOR ANALYSIS

Factor analysis is useful when dealing with multivariate analysis. The analysis is a statistical technique that tries to "explain a set of data in terms of a smaller number of dimensions" (Der & Everitt, 2002). More specifically, the analysis tries to identify a smaller number of influential variables in the data set in order to reduce the number of variables to a parsimonious set (Lorr, 1983). That is, the analysis tries to reduce number of variables, *p*, to a smaller set of variables. Since the questionnaire for the proposed study asks survey participants about numerous variables, there are possibilities for the data to have a redundant, correlated and complex structure. Thus, a manageable subset must be identified. Factor analysis will reduce the numerous variables to a smaller number of influential variables ordered by factor loadings. For example, if 50 variables are identified from the survey, it is computationally reasonable to create clusters based on a reduced number of factors, rather than based on all 50 variables. The basic structure of factor analysis follows the form of linear multiple regression.

The simple description of the model is as follows:

$$
\begin{aligned}
x_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{ik} f_k + \mu_1 \\
x_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2k} f_k + \mu_2 \\
&\vdots \\
x_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pk} f_k + \mu_p
\end{aligned}
\tag{C.1}
$$

$Where,$

$x_i = dependent\ \mathrm{var}iables$

$\lambda_{ij} = factor\, loading\ \ for\ the\ \ j th\ \ factor\ \ for\ \mathrm{var}iable\ i$

$f_i = common\ factors$

$\mu_i = residual\, terms$

As a matrix from, the above model will be denoted:

$$
x = \Lambda f + u
\tag{C.2}
$$

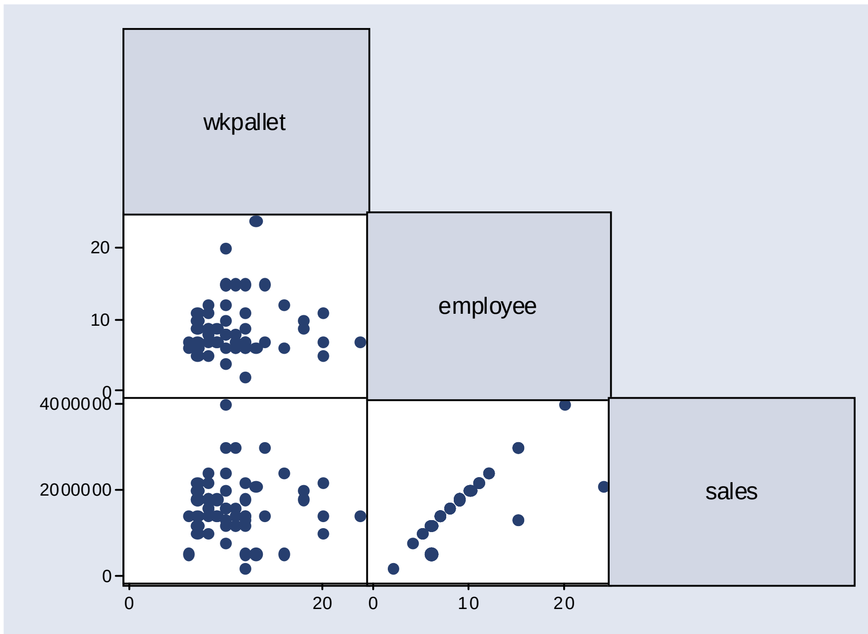
$where,$

$$
\Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{1p} & \cdots & \lambda_{pk} \end{bmatrix},\ f = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix}
$$

The residual terms $u_1, \cdots, u_p$ are assumed uncorrelated with each other and with the common factors. The elements of Λ are referred to as factor loadings. Because factors are unobserved, their locations and scales are arbitrarily fixed, if the factors are standardized with mean zero and standard deviation one.

In this analysis, factor loadings, λ , that are close to one are sought. This suggests that a variable, x , is largely influence by common factors, explaining the impacts on a dependent variable well. That is, variables with high factor loadings are thought to be highly influential in describing the factor, while variables with low factor loadings are thought to be less influential to the dependent variable.

APPENDIX D. CORRELATION PLOTS: WKPALLET, EMPLOYEE, AND SALES



APPENDIX E. POISSON REGRESSION

When the data is count, the application of linear regression often results in "inefficient, inconsistent, and biased estimates. (Long and Freese, 2003). Ordinary least square (OLS) may mislead the result. OLS treats the count data as continuous variable; however, the count data is non-negative integer (Long, 1997). Count data indicate the number of times the events have happened. The application of count data is frequently modeled in transportation studies (Washington, Karlaftis, and Mannering, 2003). For example, the number of cars arrived at the toll boots in a given period can be modeled using a Poisson distribution.

The Poisson distribution has a general form:

$$\Pr(y \mid \mu) = \frac{e^{-\mu}\mu^{y}}{y!} \quad \mathit{for}\ y = 0,1,2,\cdots \qquad \text{(E.1)}$$

where,

μ = the expected number of events

y = the number of events occurred.

The first purpose of the Poisson regression model is to estimate a μ - the rate of occurrence or the expected number of times and event will occur over a given period of time. A μ has four properties

(1) A μ u is the mean of the distribution. As a μ increase, the mass of the distribution shifts to the right.

(2) A μ is also the variance. In real data, many count variables have greater than the mean, which is called over-dispersion.

(3) As a μ increases, the probability of a zero count decreases.

(4) As a μ increases, the Poisson distribution approximates a normal distribution.

The following is the equation to estimate a μ

$$\mu_i = E\left(y_i \mid x_i\right) = \exp(x_i \beta_i) \qquad \text{(E.2)}$$

where,

y_i = event

x_i = independent variables

β_i = coefficients

Estimated parameters are used to make inference on population or predict the number of events. Maximum likelihood estimates produced Poisson parameters that are consistent, asymptotically normal, and asymptotically efficient.

APPENDIX F. POISSON REGRESSION ANALYSIS AND NEGATIVE BINOMIAL REGRESSION ANALYSIS

The number of trips is non-negative integer that can only take the values of one or two. In other words, the number of trips cannot be considered as a continuous variable. Thus, it is worthwhile considering the use of Poisson regression analysis, which is widely used for count data modeling.

In order to see if Poisson regression would predict the number of trips well, the number of trips is regressed on *employees, sales, size dummy, location dummy* and *store types dummy*. As seen in Table E-1, the model does not fit well. Moreover, Figure E-1 compares the observed and predicted shares for the stores with one and two delivery trips, respectively. The observed shares of the one-trip and two-trip stores are about76 percent and 24 percent, respectively while the predicted probabilities are only about 35 and 21 percent, respectively. Since Poisson distribution includes zero count, the model tries to predict the number of zero trips that is nonexistent in the furniture chain data set. This causes under-prediction of one-trip stores.

Negative binomial regression is presented in the second column of Table F-1 to account for the over-dispersion problem of Poisson model, which assumes the variance is equal to the mean. The over-dispersion is a common problem with the Poisson model that occurs when the variance is much larger than the mean of the variables. The over-dispersion of Poisson model makes the model difficult to be fitted to the real-world data (Long, 1997). That is, the Poisson models tend to underestimates the amount of dispersion in the outcome. The negative binomial regression model introduces a new parameter, alpha, to reflect unobserved heterogeneity among observations.

The result is similar to that of the Poisson regression. More importantly, the bottom of the table shows the likelihood-ratio test of an additional parameter, alpha. The null hypothesis is that the additional parameter does not improve the model. In other words, there is no over-dispersion problem in observed data set. The hypothesis test shows the null hypothesis cannot be rejected, concluding that negative binomial regression model is not the proper technique for Furniture Chain A data, too. Even though In conclusion, both Poisson and negative binomial regression models are not appropriate for the data set.

Table F-1. Poisson vs. Negative Binomial Models

Variables	Poisson	Negative Binomial
trip		
employee	-0.0158	-0.0158
sales	1.50E-07	1.50E-07
size2	0.0402	0.0402
size3	0.0591	0.0591
mall1	0.0743	0.0743
stchar1	0.5646	0.5646
stchar3	0.1118	0.1118
constant	-0.0692	-0.0692
lnalpha Constant		-27.4337

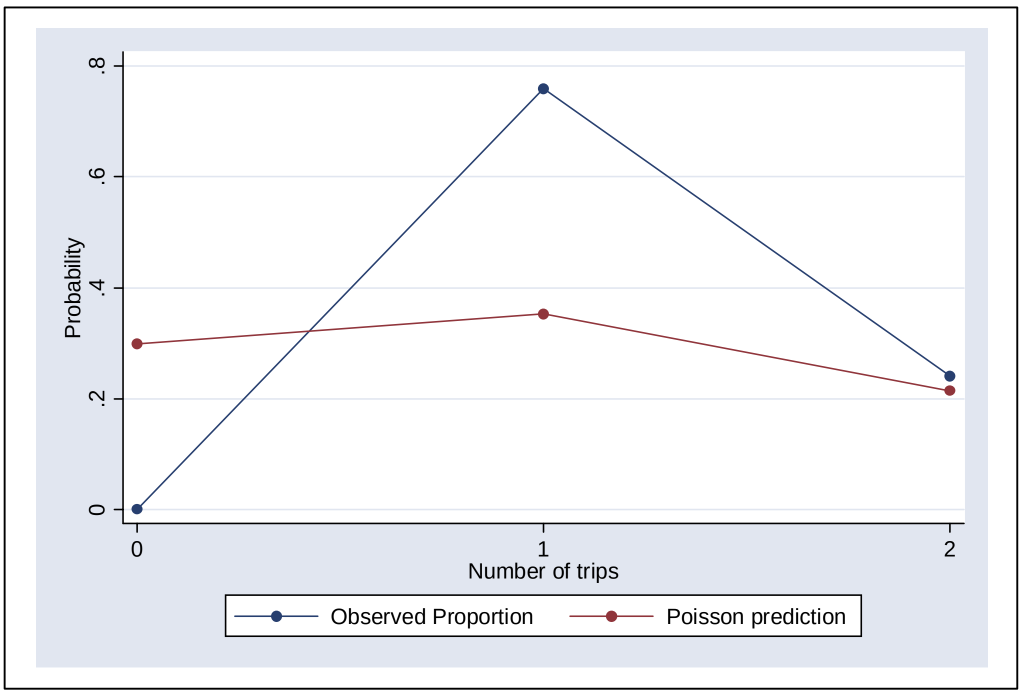


FIGURE F-1. PREDICTED PROBABILITY BY POISSION REGRESSION

APPENDIX G. TTG ESTIMATION USING ITE TRIP GENERATION METHOD

ITE Trip Generation (ITE, 2003) provides average trip rates for different types of land use. Typically, rates are calculated with respect to more than one characteristics of the target land use. For furniture stores, the rates are provided based on gross floor area (per 1000 sqft) and also employees (per employee). For each type, a total of nine models, covering different times of day and days of week are presented.

Since the analysis time period for this study is one week, three models: weekday, Saturday, and Sunday, were used to estimate the number of truck trip ends per week. Also, since exact floor area was not available for Furniture Chain A and E stores, number of employees was used as the independent variable to estimate the number of trip ends.

It should be noted that the ITE manual does not provide separate model for trucks and cars. Rather, it states *"truck trips accounted for approximately 1 to 13 percent of the weekday traffic at the sites surveyed. The average for the sites that were surveyed was approximately 5 percent. (ITE 2003, Volume 3 p. 1648)."* Thus, for calculating the number of truck trip ends, it was assumed that 5 percent of the total trip ends given by the ITE models were trucks.

The average number of trip ends (including trucks and cars) per employee for weekday, Saturday and Sunday are: 12.19, 13.87, and 12.97, respectively. Assuming that 5 percent of total trips ends are trucks, the weekly truck trip ends per employee can be calculated as

Truck trip ends per week per employee = (12.19 x 5 + 13.87 +12.97) x 0.05 = 4.3895. (G-1)

The total number of employees for all 58 stores of Furniture Chain A is 516. Thus, the total truck trip ends for Furniture Chain A is estimated as

Total truck trip ends per week for Furniture Chain A = 4.3895 x 516 = 2265 (G-2)

Similarly, the truck trip ends for Furniture Chain E, which employs a total of 134 employees in 11 stores, can be estimated as:

Total truck trip ends per week for Furniture Chain A = 4.3895 x 134 = 588 (G-3)

References

Al-Deek, H.M., Johnson, G., Mohamed, A. and El-Maghraby, A. (2000). Truck Trip Generation Models for Seaports with Container and Trailer Operation. Transportation Research Record 1719, 1-9.

Al-Deek, H.M. (2001). "Comparison of Two Approaches for Modeling Freight Movements at Seaports" The ASCE Journal of Computing in Civil Engineering. Vol. 15, No. 4. PP. 309-319.

Boerkamps, J.H.K., van Binsbergen, A.J. and Bovy, P.H.L. (2000). Modeling Behavioral Aspects of Urban Freight Movement in Supply Chains. Transportation Research Record 1725, 17-25.

Bowersox, D.J., Closs, D.J. and Cooper, M.B. (2002). Supply Chain Logistics Management. Boston, MA: McGraw Hill.

Brogan, J.D. (1979). Development of Truck Trip-Generation Rates by Generalized-Land Use Categories. Transportation Research Record 716, 38-43.

Brogan, J.D. (1980). Improving Truck Trip-Generation Techniques through Trip-End Stratification. Transportation Research Record 771, 1-6.

BTS (2004). Freight Shipments in America: Preliminary Highlights from the 2002 Commodity Flow Survey Plus Additional Data. U.S. Bureau of Transportation Statistics. Washington, D.C.; U.S. Department of Transportation.

BTS. (2005). 2001 National Household Travel Survey: Introduction. Available at http://nhts.ornl.gov/2001/html_files/introduction.shtml. U.S. Bureau of Transportation Statistics. Washington, D.C.; U.S. Department of Transportation. (April 27, 2005)

Buchheit, R.B. (2002). Vacuum: Automated Procedures for Assessing and Cleansing Civil Infrastructure Data. Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Cambridge Systematics, Inc. (1997). A Guidebook for Forecasting Freight Transportation Demand (NCHRP Report 388). Washington, D.C.: TRB. National Research Council.

Child, D. (1990). The Essentials of Factor Analysis. Second Edition. London, UK: Cassell Educational Limited.

Crosier, S. (2004). Geocoding in ArcGIS. Redlands, CA: ESRI

Der, G. and Everitt, B.S. (2002). A Handbook of Statistical Analyses Using SAS, Second Edition. New York: Chapman and Hall/CRC.

Disney, S.M., Potter, A.T. and Gardner, B.M. (2003) The Impact of Vendor Managed Inventory on Transport Operations. Transportation Research, Part E, 39, 363-380.

ESRI. (2004). StreetMap USA. Redlands, CA: ESRI.

FHWA. (2002). Freight Analysis Framework. Freight News. Federal Highway Administration. Washington D.C.

FHWA (2004[a]) The Freight Story. Federal Highway Administration. Available at <www.ops.fhwa.dot.gov/freight/publications/freight%20story/freight.pdf> (February 12, 2004).

FHWA. (2004[b]) Freight Analysis Framework. Federal Highway Administration, Available at <http://www.ops.fhwa.dot.gov/freight/adfrmwrk/index.htm> (March 29, 2004).

Fisher, M.J. and Han, M. (2001) Truck Trip Generation Data (NCHRP Report 298). Washington, D.C: TRB, National Research Council.

Fite, J.D., Taylor, G.D., Usher, J.S., English, J.R. and Roberts, J.N. (2002). Forecasting Freight Demand Using Economic Indices. International Journal of Physical Distribution and Logistics Management, 32 (4), 299-308.

Gaither, N. and Frazier, G. (1999). Production and Operations Management. Eighth Edition. Cincinnati, OH.: South-Western College Publishing.

Garrido, R.A. (2001, August). Insights on Freight and Commercial Vehicle Data Needs. Paper presented at International Conference on Transport Survey Quality and Innovation, Kruger National Park, South Africa. <http://www.its.usyd.edu.au/conferences/international_conference_on_transport_survey_quality_and_innovation%20(new)/workshop_papers.htm> (2004, April 14).

Gujarati, D.N. (2003). Basic Econometrics, Fourth Edition. Boston, MA: McGraw-Hill.

Holguín-Veras, J. and López-Genao, Y. (2002). Truck Trip Generation at Container Terminals: Results from a Nationwide Survey. Transportation Research Record 1790, 89-96.

InfoUSA. (2005) InfoUSA. Available at http://lists.infousa.com/images/ir/ir_index.html

(May 19, 2005)

Iding, M.H.E., W. J. Meester, and L.A. Tavasszy. (2002). Freight Trip Generation by Firms. Paper presented at the 42nd European Congress of the Regional Science Association. Dortmund, Germany.

ITE (2003). *Trip Generation*, 7th *Edition. Volume 1 & 3*. Institute of Transportation Engineers. Washington, D.C.

ITE (2003b). *Trip Generation Handbook: An ITE Recommended Practice*. Institute of Transportation Engineers. Washington, D.C. pp.106

Klodzinski, J. and Al-Deek, H.M. (2003). Transferability of an Intermodal Freight Transportation Forecasting Model to Major Florida Seaports. Transportation Research Record 1820, 36-45.

Khouha, M. (2003). Optimizing Inventory Decisions in a Multi-stage Multi-customer Supply Chain. Transportation Research, Part E, 39,193-208.

List, G.F. and Trunquist, M.A. (1994) Estimating Truck Travel Patterns in Urban Areas. Transportation Research Record 1430, 1-9.

Long, J.S. (1997). Regression Analysis Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.

Long, J.S. and Freese, J. (2003). Regression Models for Categorical Dependent Variables using STATA. Revised Version. College Station, TX: Stata Press.

Lorr, M. (1983). Cluster Analysis for Social Scientists. San Francisco, CA.: Jossey-Bass Publishers.

McCarthy, P.S. (2001). Transportation Economics: Theory and Practice – A Case Study Approach. Malden, MA: Blackwell Publishers, Inc.

Mentzer, J.T., DeWitt, W., Keebler, J.S., Min, S., Nix, N.W., Smith, C.D. and Zacharia, Z.G. (2001). Defining Supply Chain Management. Journal of Business Logistics, 22 (2), 1-25.

Miller, T. and de Matta, R. (2003). Integrating Production and Transportation Scheduling Decision between Two Geographically Separated Plants. Journal of Business Logistics, 24 (1), 111-146.

Miranda, P.A. and Garrido, R.A. (2004). Incorporating Inventory Control Decisions into a Strategic Distribution Network Design Model with Stochastic Demand. Transportation Research, Part E, 40, 183-207.

Nagurney, A., Dong, J., and Zhang, D. (2002) A Supply Chain Network Equilibrium Model. Transportation Research, Part E, 38, 281-303.

Pendyala, R.M., Shankar, V.N. and McCullough, R.G. (2000). Freight Travel Demand Modeling: Synthesis of Approaches and Development of a Framework. Transportation Research Record 1725, 9-16.

Pipkin. J.S. (1995). "Disaggregate Models of Travel Behavior," in The Geography of Urban Transportation. Ed. (Susan Hanson). New York: The Guilford Press

Sen, A. and Srivastava, M. (1990). Regression Analysis: Theory, Methods, and Applications. New York: Springer.

Slavin, H.L. (1974). Demand for Urban Goods Vehicle Trips. Transportation Research Record 591, 32-37.

Suarez-Villa, L. (2003). The E-economy and the Rise of Technocapitalism: Networks, Firms, and Transportation. Growth and Change, 34 (4), 390-414.

Tadi, RR & Balbach, P. (1994) "Truck Trip Generation Characteristics of Nonresidential.

Land Uses". ITE Journal, 64 (7).

TRB. (2002). Regulating Weights, Lengths, and Widths of Commercial Motor Vehicles (Special Report 267). Transportation Research Board. Washington, D.C.

U.S. Bureau of Census. (2005). Your Gateway to Census 2000. Available at http://www.census.gov/main/www/cen2000.html. Washington, D.C.; U.S. Department of Commerce. (August 25, 2005)

U.S.Bureau of Census. (1997). 1997 Commodity Flow Survey (CD-EC97-CFS). Washington, D.C.; U.S. Department of Commerce.

U.S. Congress. (1991). Public Law, no. 240, 102th Cong., December 18, 1991.

U.S. Congress. (1998). Public Law, no. 178, 105th Cong., June 9, 1998

U.S. DOT. (2004). 2001 National Household Travel Survey. U.S. Department of Transportation. Washington, D.C.:

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2003). Statistical and Econometric Methods for Transportation Data Analysis. New York: Chapman and Hall/CRC.

Winston, W.L. (1994). Operations Research, Third edition. Belmont, CA: Duxbury Press.

Wisner, J.D. (2003). A Structural Equation Model of Supply Chain Management Strategies and Firm Performance. Journal of Business Logistics, 24 (1), 1-26.