

2013

## Investigation of Truckload Prices in the U.S: Exploratory Analysis, Forecast Methods, and the Influences of Unemployment on Freight Prices

Alex M. Sanchez  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Sanchez, Alex M., "Investigation of Truckload Prices in the U.S: Exploratory Analysis, Forecast Methods, and the Influences of Unemployment on Freight Prices" (2013). *Graduate Theses, Dissertations, and Problem Reports*. 4999.

<https://researchrepository.wvu.edu/etd/4999>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

**Investigation of Truckload Prices in the U.S: Exploratory Analysis, Forecast Methods, and  
the Influences of Unemployment on Freight Prices.**

**Alex M. Sanchez**

Dissertation submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy  
in  
Engineering

Avinash Unnikrishnan, Ph.D. Chair

David R. Martinelli, Ph.D.

Udaya B. Halabe, Ph.D.

Donald Lacombe, Ph.D.

Santiago Pinto, Ph.D.

Department of Civil and Environmental Engineering

Morgantown, West Virginia

2013

Keywords: Truckload Prices; Freight Transportation; Forecasting; Spatial Econometric; Time Series; Exploratory Spatial Data Analysis

**Copyright 2013 Alex M. Sanchez**

## **ABSTRACT**

### **Investigation of Truckload Prices: Exploratory Analysis, Forecast Methods, and the Influences of Unemployment on Freight Prices.**

**Alex M. Sanchez**

Truckload (TL) pricing is a major factor that influences the manufacturing and retail costs of products. In the U.S., trucks accounts for more than 90% of freight shipped based on value, and it is expected to grow in the following years. TL price setting is a very complex task for logistic companies as it depends on a number of factors including the logistics carriers' business strategies and other social and economic variables. Understanding TL patterns across the U.S. is important not only for logistic companies, but also for policy makers. TL prices are commonly provided on a dollar per mile rate. Thus the total transportation costs on a route will be the product of the truckload price rate and the distance. More accurate prediction of TL price will enable logistic companies to develop more optimal strategies to operate their transportation activity across destinations and effectively allocate resources on potential demand locations. Freight and economic policy makers will also be able to use this information to explore different potential economic scenarios.

This research analyses private data sets (TL rates), and publicly available data such as diesel cost, unemployment, wages, population, and gross state product to understand trends in TL prices. TL rates are evaluated through exploratory and visualization techniques to obtain useful insights. Time series analysis (TSA) and spatial econometric analysis (SEA) are conducted for forecasting TL prices. TSA provides with a general model based on time and delivery distance between origin and destination. Spatial econometric panel models incorporate the spatial dependency, being used for drawing inferences across space, and also for forecasting TL prices. Results indicate that TL prices are closely associated with unemployment, which links the consumer spending with transportation cost. Diesel cost has not impacted TL prices significantly during the last years, as is evidenced in the TSA and SEA. Moreover, in low demand condition such as high unemployment, carriers are likely to serve larger delivery distance in order to reduce TL prices, which impact TL prices in neighboring locations. Increasing the delivery distance by 1.00% was found to reduce the price in dollar-per-mile by about -0.25%, and raise prices in neighboring locations by about +0.05%. Similarly, 1% increase in unemployment rate was found to reduce prices by about -0.30% and increase prices in neighboring locations by about +0.06%. Forecasting models indicate accurate TL price values, with MAPE values less than 10% for the TSA model for estimating an overall monthly price in the U.S.; and less than 20% for the SEA that consider spatial dependence for estimating a yearly price at each U.S. state. This research represents a benchmark in the analysis of freight prices, providing useful insights, identifying significant variables impacting TL prices, and potential methodologies for forecasting truckload prices.

## **DEDICATION**

I would like to dedicate this work to my father Wilder, my mother Graciela, my siblings Carlota, Raul, and Ivan, and my grandmother Celia (rests in peace), who always have expressed their love and support. Finally, dedicate to the new members of my family, my nieces and nephews, Ximena, Claudia, Raul and Francisco.



## **ACKNOWLEDGEMENTS**

I would like to express my thanks and gratitude to my advisor and committee, Dr. Avinash Unnikrishnan, Dr. David Martinelli, Dr. Udaya Halabe, Dr. Donald Lacombe, and Dr. Santiago Pinto for their comments and guidance throughout this journey. I want to acknowledge and thank Avinash Unnikrishnan, my advisor, for his guidance, insightful comments, and for this research opportunity. This work would have not been possible without the help, support and advice from my committee.

Also, I would like to thank Varunraj Valsaraj of ArrowStream Logistics for providing the data sets that makes this work possible. I would also like to thank the Mid-Atlantic Universities Transportation Center (MAUTC) for supporting my education at West Virginia University.

## TABLE OF CONTENTS

<b>ABSTRACT.....</b>	<b>II</b>
<b>DEDICATION.....</b>	<b>III</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>IV</b>
<b>1.0 INTRODUCTION.....</b>	<b>1</b>
1.1 Motivation .....	1
1.2 Transportation Industry .....	3
1.3 Transportation and Geography.....	6
1.4 Problem Statement .....	7
1.5 Research Objectives .....	8
1.6 Organization of the Research .....	9
<b>2.0 LITERATURE REVIEW .....</b>	<b>10</b>
2.1 Introduction .....	10
2.2 Exploratory Data Analysis (EDA) .....	10
2.2.1 Robust Statistics .....	11
2.2.2 The Smooth and Rough Part of the Data.....	13
2.3 Time Series Analysis.....	13
2.4 Exploratory Spatial Data Analysis (ESDA).....	15
2.4.1 Geographic Information System (GIS) .....	17
2.5 Data Visualization.....	18
2.5.1 Data Mapping.....	18
2.5.2 Representing Interaction between Origin-Destination Trips .....	18
2.5.3 Spatial Autocorrelation .....	21
2.6 Spatial Econometric Analysis .....	22
2.7 Spillover Impacts .....	26
2.8 Forecasting with Spatial Econometric .....	28
<b>3.0 METHODOLOGY .....</b>	<b>30</b>
3.1 Introduction .....	30
3.2 Time-Series Analysis .....	31

3.2.1 Autoregressive and Moving Average Process .....	31
3.2.2 Autocorrelation Data Diagnostic.....	32
3.2.3 Forecasting Values Procedure .....	33
3.3 Smoothing Techniques.....	34
3.4 Model Validation and Accuracy .....	35
3.5 Exploratory Spatial Data Analysis (ESDA).....	36
3.5.1 Spatial Weight Matrix .....	36
3.5.2 Measures of Spatial Autocorrelation.....	38
3.6 Spatial Econometric Models .....	39
3.6.1 Spatial Panel Models.....	41
3.7 Spatial Test for Spatial Dependency and Effects.....	43
3.7.1 The Lagrange Multiplier Test .....	43
3.7.2 Random vs. Fixed Effects (Hausman Test).....	45
3.8 Spatial Spillover Impacts .....	46
<b>4.0 EXPLORATORY DATA ANALYSIS (EDA).....</b>	<b>48</b>
4.1 Introduction .....	48
4.2 Descriptive Statistics.....	49
4.3 EDA Plots .....	51
4.4 Data Diagnostic Plots.....	57
4.5 Discussion and Conclusion .....	59
<b>5.0 TIME SERIES ANALYSIS .....</b>	<b>61</b>
5.1 Autoregressive Model (AR), Model Calibration and Estimation .....	61
5.2 Smoothing Process.....	62
5.3 Results .....	63
5.4 Performance Evaluation of Methodologies.....	65
5.5 Conclusion .....	66
<b>6.0 EXPLORATORY SPATIAL DATA ANALYSIS (ESDA) .....</b>	<b>67</b>
6.1 Introduction.....	67
6.2 Exploratory Maps.....	68
6.2.1 Dry Loads .....	68
6.2.2 Reefer Loads .....	74

6.3 Interaction Maps.....	79
6.4 Spatial Autocorrelation .....	84
6.5 Conclusion .....	90
<b>7.0 SPATIAL ANALYSIS .....</b>	<b>94</b>
7.1 Introduction .....	94
7.2 Database .....	95
7.3 Spatial Weight Matrix .....	96
7.4 Spatial Autocorrelation .....	96
7.4.1 The Lagrange Multiplier Test .....	97
7.5 Spatial Econometric Model Specification.....	98
7.5.1 Spatial Autoregressive Model (SAR).....	99
7.5.2 Fixed v.s. Random Effects (Hausman Test).....	99
7.5.3 Parameter Estimation .....	101
7.5.4 Forecasting Results .....	107
7.6 Summary and Conclusions.....	107
<b>8.0 CONCLUSION AND FUTURE WORK .....</b>	<b>109</b>
8.1 Findings.....	109
8.2 Discussion .....	112
8.3 Contribution of Dissertation.....	112
8.4 Limitations and Recommendation for Future Research.....	114
<b>REFERENCES.....</b>	<b>117</b>
<b>APPENDIX A: EXPLORATORY MAPS .....</b>	<b>125</b>
A.1. Dry Load Managed Lanes .....	125
A.2 Dry Loads non-Managed Lanes .....	128
A.3 Reefer Loads - Lane Managed .....	131
A.4. Reefer Load non-Managed Lanes .....	134
A.5 U.S. Unemployment Maps .....	137
A.6 U.S. Unemployment vs. TL Prices.....	138
<b>APPENDIX B: SPATIAL PANEL SAR AND SDM .....</b>	<b>139</b>
B.1 SAR Fixed Effects .....	139
B.2 SAR Random Effects .....	141

B.3 SDM Fixed Effects .....	143
B.4 SDM Random Effects .....	145
<b>APPENDIX C: SPATIAL PANEL SEM AND SDEM.....</b>	<b>147</b>
C.1 SEM Fixed Effects .....	147
C.2 SEM Random Effects .....	149
C.3 SDEM Fixed Effects .....	151
C.4 SDEM Random Effects .....	153

## LIST OF TABLES

TABLE 1: DATA SAMPLE.....	49
TABLE 2: DATASET DESCRIPTIVE STATISTICS .....	50
TABLE 3: MODEL A RESULTS COMPARISON .....	64
TABLE 4: MODEL B RESULTS COMPARISON.....	65
TABLE 5: LAGRANGE MULTIPLIER TEST FOR SPATIAL PANEL DATA .....	97
TABLE 6: HAUSMAN TEST BETWEEN FIXED AND RANDOM EFFECTS .....	100
TABLE 7: SPATIAL PANEL AUTOREGRESSIVE MODELS (SAR_F) WITH SPATIAL FIXED EFFECTS .....	102
TABLE 8: SPATIAL PANEL AUTOREGRESSIVE MODELS (SAR_R) WITH SPATIAL RANDOM EFFECTS.....	104
TABLE 9: SPATIAL ERROR MODELS (SEM_F) WITH SPATIAL FIXED EFFECTS .....	105
TABLE 10: SPATIAL ERROR MODELS (SEM_R) WITH RANDOM EFFECTS .....	106
TABLE 11: ACCURACY OF STATE LEVEL TL PRICE FORECASTING .....	107

## LIST OF FIGURES

FIGURE 1: REGION'S LOCATIONS .....	37
FIGURE 2: ROW-STANDARDIZED FIRST-ORDER CONTIGUITY MATRIX .....	37
FIGURE 3: SCATTER PLOT- PRICE VS. DISTANCE .....	51
FIGURE 4: DISTRIBUTION DIAGRAM AND BOXPLOT OF PRICE-PER-MILE.....	53
FIGURE 5: PRICE FLUCTUATION OVER TIME .....	54
FIGURE 6: MULTIVARIATE SCATTER PLOT FOR DRY MANAGED .....	55
FIGURE 7: STEAM-AND-LEAF PLOT .....	56
FIGURE 8: PRICE-PER-MILE VS. DISTANCE PLOTS .....	56
FIGURE 9: AUTOCORRELATION PLOTS.....	58
FIGURE 10: SMOOTHING TECHNIQUES .....	62
FIGURE 11: DRY LOAD - SIX YEARS AVERAGED .....	68
FIGURE 12: DRY LOAD MANAGED & NON-MANAGED BY YEAR.....	70
FIGURE 13: AVERAGED TL PRICE FOR DRY LOAD MANAGE.....	71
FIGURE 14: AVERAGE TL PRICE FOR DRY LOAD MANAGED BY O-D STATE.....	72
FIGURE 15: REEFER LOAD MANAGED - SIX YEARS AVERAGED.....	74

FIGURE 16: REEFER LOAD MANAGED BY YEAR.....	76
FIGURE 17: AVERAGED TL PRICE FOR REEFER LOAD BY DISTANCE.....	77
FIGURE 18: AVERAGE TL PRICE FOR REEFER LOAD MANAGED BY O-D STATE.....	78
FIGURE 19: FLOW MAP FOR TOTAL NUMBER OF DELIVERIES.....	80
FIGURE 20: BUNDLE MAP FOR TOTAL DELIVERIES IN THE U.S. ....	81
FIGURE 21: FLOW AND EXPLORATORY MAP .....	83
FIGURE 22: MORAN'S I SCATTER PLOT FOR EACH DATASET .....	85
FIGURE 25: U.S. REGIONAL MAP ASSOCIATED WITH TL PRICES .....	93
FIGURE A.1.1 DRY LOAD MANAGED - SIX YEARS AVERAGED .....	125
FIGURE A.1.2: DRY LOAD MANAGED BY YEAR .....	125
FIGURE A.1.3: AVERAGED TL PRICE FOR DRY LOAD MANAGED BY DISTANCE .....	126
FIGURE A.1.4: AVERAGE TL PRICE FOR DRY LOAD MANAGED BY O-D STATE .....	126
FIGURE A.1.5: FLOW MAP OVER THE YEARS: DRY LOAD MANAGED .....	127
FIGURE A.1.6: FLOW MAP - TOTAL: DRY LOAD MANAGED .....	127
FIGURE A.2.1: DRY LOAD NON-MANAGED - SIX YEARS AVERAGED .....	128
FIGURE A.2.2: DRY LOAD NON-MANAGED BY YEAR .....	128
FIGURE A.2.3: AVERAGED TL PRICE FOR DRY LOAD NON-MANAGED BY DISTANCE .....	129
FIGURE A.2.4: AVERAGE TL PRICE FOR DRY LOAD NON-MANAGED BY O-D STATE .....	129
FIGURE A.2.5: FLOW MAP OVER YEARS: DRY LOAD NON-MANAGED .....	130
FIGURE A.2.6: FLOW MAP - TOTAL: DRY LOAD NON-MANAGED.....	130
FIGURE A.3.1: REEFER LOAD MANAGED - SIX YEARS AVERAGED.....	131
FIGURE A.3.2: REEFER LOAD MANAGED BY YEAR.....	131
FIGURE A.3.3: AVERAGED TL PRICE FOR REEFER LOAD MANAGED BY DISTANCE.....	132
FIGURE A.3.4: AVERAGE TL PRICE FOR REEFER LOAD MANAGED BY O-D STATE.....	132
FIGURE A.3.5: FLOW MAP OVER YEARS: REEFER LOAD MANAGED.....	133
FIGURE A.3.6: FLOW MAP - TOTAL: REEFER LOAD MANAGED .....	133
FIGURE A.4.1: REEFER LOAD NON-MANAGED - SIX YEARS AVERAGED.....	134
FIGURE A.4.2: REEFER LOAD NON-MANAGED BY YEAR.....	134
FIGURE A.4.3: AVERAGED TL PRICE FOR REEFER LOAD NON-MANAGED BY DISTANCE.....	135
FIGURE A.4.4 AVERAGE TL PRICE FOR REEFER LOAD NON-MANAGED BY O-D STATE.....	135
FIGURE A.4.5: FLOW MAP OVER YEARS: REEFER LOAD NON-MANAGED.....	136
FIGURE A.4.6: FLOW MAP - TOTAL: REEFER LOAD NON-MANAGED .....	136
FIGURE A.5: U.S. UNEMPLOYMENT FROM 2005 TO 2010.....	137
FIGURE A.6: U.S. UNEMPLOYMENT V.S TL PRICES.....	138

# CHAPTER 1

## 1.0 INTRODUCTION

### 1.1 Motivation

Truckload (TL) pricing is a major factor that influences the manufacturing and retail costs of products and affects the effectiveness and efficiency of freight transportation. In the U.S., trucks and rail account for 65% of the nation's domestic freight volume. Both have shown significant growth over the past 10 years. In 2005, U.S highways carried 77% of the freight based on tons shipped, and 92% of freight shipped based on value. “It has been said that the rail lines and highways across the country have become the nation’s largest warehouse” (AASHTO, 2007). According to the Freight Analysis Framework (FAF) the movement of goods in term of tonnage has increased by about 13.6 percent from 2009 to 2010. From 2010 to 2040 is expected to increase at a rate of 1.1 percent per year (FHWA, 2011). This increase in freight volume has affected the truck volumes on roads, generating traffic congestion and uncertainty in delivery times. Moreover globalization and lean supply chain practices like “just-in-time delivery” and “collaborative logistics” have increased the demands on the freight systems. Over the past years, freight cost within US has varied significantly due different actors affecting carriers, e.g. consumer spending, and demand.

These freight cost variations have created large distress in a low margin business like groceries, and restaurants. Consumers are sensitive to price changes. Frequent changes in prices usually result in a drop of consumer demand since transportation links producer and consumer through production costs. It is a hard challenge to pass the increasing freight cost to end consumers without losing the consumer demand. Freight cost is an integral part of the price of any commodity. Empirical evidence indicates that transport cost accounts for 10% of the total

cost of a product, and the impact of increasing transport costs by 10% would reduce trade volumes by more than 20% (Rodrigue, 2009). The truckload industry is a very competitive market affected by various factors such as fleet availability, transportation connectivity, infrastructure condition, etc. Contracts between TL logistic companies and retail business are under a rate (\$/mile) agreement, which may not represent exactly the real transport costs, and therefore these rates could create loss or profit for the carrier (Rodrigue, 2009). Under this circumstance, setting TL rate is a complex task for logistic companies, because it depends on external factors such as fuel cost, negotiation tactics, and also on the spatial distribution of demand that depends on the market size of the freight company. Logistic companies have price strategies focused on improving efficiency, reducing costs, and expanding their market size. Other aspects that influence TL rates is the type of goods being transported. Dry and refrigerated products are shipping in vehicles with similar characteristics. However, refrigerated vehicles required air conditioning unit to control the inside temperature of the trailer to allow deliveries of temperature-sensitive products. Reefer loads have a large variety of products e.g. wine, beer, nutritional supplements, pharmaceuticals , dairy products, meats, deli foods, fresh fruits, seafood and vegetables, which usually need extra attention in the route. Dry loads vary from steel products to building materials, e.g. machinery, motor vehicles, sand, fertilizer, and road materials. Therefore, dry loads are expected to have lower transportation costs than reefer loads.

The understanding of freight patterns such as movement activity, which allows for identifying locations under high and low demand of products, and the TL price behavior across locations are useful information for logistic companies and allow them to more accurately plan their operations across destinations. It helps logistic companies to effectively allocate resources on potential high demand locations. Developing accurate TL price forecasting models will allow



the freight operators to negotiate better contracts which more accurately reflect the actual transportation costs.

A few studies related to carrier pricing has been conducted; however, the underlying structure of the TL rates and potential variables that affects them has not been established. Methodologies such as regression analysis has been used for estimating freight rates, in both truckload (TL) transportation (Swenseth and Godfrey 1996) and in less-than-truckload (LTL) transportation (Hall 1985; Kay et al., 2009; Ozkaya et. al. 2010; Swenseth and Godfrey 2001). Results from these studies indicated forecast values overestimate the actual truckload prices in a range between fourteen to thirty seven percent. Therefore, the development of new methods that can provide more accurate predictions is required. Studies that associates consumer spending with socioeconomic variables such as population, gross state product, fuel cost and unemployment are found in the literature (Thaiprasert, 2011). For instance, high unemployment rates reduce consumer demand of some commodities, and high fuel cost tends to reduce people's demand in non-essential services or activities. To date, the relationship between truckload prices and the economy of a region is still not well established, mainly due to the scarceness of data available. This research is an original study that provides important insights and modeling techniques for the TL market in the U.S.

## **1.2 Transportation Industry**

In the U.S. the freight transportation industry is a competitive business that serves the domestic and the international market through a supply chain network with various modes of transportation such as air, maritime, rail, and truck. This industry moved a total of \$1.1 trillion in 2009, which represented an average of 9.4 percent of the annual gross domestic product (GDP)

between 2000 and 2008 (SelectUSA, 2012). In 2009 motor carriers have moved 8.8 billion tons of freight, which represents the 68 percent of all domestic freight tonnage; and obtained \$544 billion of revenues, which represents the 81.9 percent of the total revenue from all domestic transport modes (ATA, 2011). Two major markets are under the trucking industry: less than truckload (LTL) and the truckload (TL). LTL provides partial-load shipments to multiple destinations with shipment's weights less than 10,000 pounds, LTL rates are based on shipment weight. TL provide trailer-full load deliveries to a single destination, limited to the available space in the trailer, characterized by shipment's weights of 10,000 pounds or more.

The trucking industry has undergone a major transformation in the 1980's due to the deregulation and with the culmination of the Interstate Highway System in 1991. After deregulation the structure of the trucking industry changed, the TL industry became more important than the LTL business. LTL is associated with unionized labor, and TL with non-unionized labor, it makes TL rates lower and more competitive than LTL carriers. In general, deregulation has decreased truck transport costs in the U.S., and has improved carriers' business, by being more efficient by reducing delivery times (RITA, 2009).

TL prices are based on a per-mile rate plus a fuel surcharge. The fuel surcharge is either cents per mile rate or a percentage of the freight bill. In addition, the TL rate also depends on the way logistics operations are managed along the route from origin to delivery point. Logistic companies coordinate operations on specific routes with or without the help of third party logistics (3PL) companies to ensure efficient and reliable movement of goods. These routes are called managed lanes. Cost of managed lanes is more reliable than non-managed lanes since the logistic company is responsible for managing the day to day operations.

Freight carriers like other business normally operate by developing pricing strategies in order to preserve their market position, increase market size, and also to reduce their costs. Price strategies such as cross-subsidization are usually common in the trucking industry. In this research, truckload price database shows this characteristic. Freight rates vary from state to state, by type of commodity, and by origin and destination of the deliveries. TL rates can be significantly different from the same origin and destination for dissimilar commodity, and same commodity with equal delivery distances but different destinations. In economic theory the concept of cross-subsidization is pricing the same service differently in different locations so that the companies can increase their market size. Church and Ware (1999) state that cross-subsidy exists "if the revenues from a product are less than its costs of production". Similarly, "price discrimination," which implies cross subsidization is mostly implemented by companies with high profits, because profits made in one market may allow companies to widen their market size by cross-subsidizing loss-making services in not economically attractive locations. In freight transportation, "product" refers to the services provided by carriers for transporting goods from origin to final destination; the price (dollar-per-mile) can be significantly different from and to particular locations.

To date there is not much literature regarding TL carriers and price strategies. It seems that the majority of technical papers were published after truck deregulation took place, in the early 1980's. Previous studies, before deregulation, attributed these variations to a non-competitive behavior. However, Beilock (1986), after deregulation, investigated this pricing variation. Results from this study determined that pricing variations are associated with a competitive market in the trucking industry. A competitive market involves for instance the willingness to pay more for faster and more reliable deliveries when the product is a high value

or perishable. Low rates can be provided when carriers filled backhauls quickly. Thus, carriers can charge differently based on the type of product and service, and also when the availability of truck units is scarce, so the carrier has to prioritize the high price deliveries. Finally, variations of TL rates are likely due to a competitive market; thus, current market conditions of supply and demand takes an important part in pricing.

### **1.3 Transportation and Geography**

The relationship of spatial geography and transportation cost is still not considered in current methods for modeling the freight transportation activity. From the engineering perspective, geography is recognized as a major issue in the design and construction of transportation infrastructure. Topography, geology and hydrology play an important role in the spatial location of the transportation infrastructure. Generally, an optimal location is the one which reduces construction costs; provides an adequate level of service and connectivity across units. Moreover, the transportation engineering field is usually under the analysis of large data sets, and mostly geo-referenced data. For instance, traffic counts, travel time, and commodity flows fall into this category. The analysis of this unstructured data creates new challenges and approaches for making these data sets understandable and for identifying potential patterns that will become useful insights. Such insights cannot be obtained from traditional analysis methods. A few technical papers are available in the literature review that describe and illustrate the link between geography, freight transportation costs, and the economy of a region. One potential reason for the relatively few papers in this area might be due to the lack of data available and constraints of publishing private data.

The major issue facing logistic companies is to improve the efficiency and reliability of the deliveries to reduce the cost of moving goods. Management techniques such as optimization of routes, and truckload collaboration are implemented for this purpose. These techniques rely on the spatial location of transportation infrastructure and accessibility between the origin and destination of shipments. Spatial location, infrastructure condition and availability, and adequate capacity to handle freight in seaports, roadways, and airports have impacted logistic costs. It is clearly noted that spatial location plays an important role in the analysis of freight transport but it is still not clear how to approach large data sets with the space dimension. Exploratory spatial data analysis helps to understand the link between the geography and transport activity, throughout visualization of data such as maps that allows users to perceive significant insights. This research provides novel approaches for displaying data sets across space, and interaction between locations, allowing for a comprehensive understanding of large data sets.

#### **1.4 Problem Statement**

The understanding of current freight patterns related to TL prices is critical for helping freight, logistic and supply chain agents in developing accurate forecasting models, reliable planning and operation management techniques to reduce transport costs. Fuel cost, shipping distance, type of commodity (dry/refrigerated), people spending, and correlation among these variables over time and space affect the TL prices in the U.S., but to-date there are no studies available in the published literature which study the impact of these factors on TL prices. Logistic companies and freight agents depend on reliable price forecasting, and in understanding the underlying structure of TL prices across locations when forming contracts. The total cost of truck route is computed as the product of the truckload price and the distance of the route.

Therefore, accurate estimates of the TL price will enable the logistic companies to determine optimal routes to reduce the transportation costs.

There is a need to understand the TL price patterns, significant factors affecting them, and the distribution of truckload movements in the U.S. There is also a need to develop a methodology which can exploit the spatial and temporal nature of the data set to generate insights about the TL price variation and provide reasonably accurate forecasts. This research is the first to use advanced exploratory spatial data analysis tools to visualize and characterize the TL price variation across the U.S. Time series and spatial analysis based regression models are developed to accurately forecast and to quantify the impact of factors such as unemployment and delivery distances on TL price.

## **1.5 Research Objectives**

This research answers the following questions:

1. Can we define an appropriate state group or cluster based on freight prices to determine regions in the U.S. which have similar patterns related to freight prices?
2. How do factors such as spatial location, fuel price at origin and destination, delivery distance, and socioeconomic factors affects the truckload price in the U.S.?
3. Does a spatial dependency and knowledge spillover in freight prices exist and can they be quantified?
4. Can we accurately forecast future truckload freight prices based on variables such as delivery distance, fuel cost and socioeconomic factors?

## **1.6 Organization of the Research**

The organization of this dissertation is as follows. Chapter 1 is the introduction of the research, where a background of the transportation industry in the U.S., and their relationship with geography is described, the problem statements and objectives are also given. This chapter is intended to provide readers with a broad picture of transportation and logistics industry basically in the truckload business. A literature follows as the subject of Chapter 2 to provide the related technical's papers on the topic as well as papers in related areas affecting truckload prices. Chapter 3 formulates the methods for formulating and modeling the models to forecast truckload prices into a mathematical form. Chapter 4 presents the exploratory data analysis (EDA), a univariate analysis of the truckload price database. Chapter 5 formulates the time series models, considering patterns found in chapter 4. Chapter 6 presents the exploratory spatial data analysis (ESDA), multiple maps are provided along with insights. Chapter 7 presents the spatial econometric analysis, where inference analysis is made between significant variables and TL price. Chapter 8 pointed out the contributions of this research, and concludes the research, providing issues for future study. The references and appendixes are attached at the end of the dissertation.

## CHAPTER 2

### 2.0 LITERATURE REVIEW

#### 2.1 Introduction

The advances in computing hardware in terms of memory and processing power and software during the last decades have allowed researchers to perform sophisticated and complex data analyses. Significant improvement in image processing capabilities of common desktop computers made visualization of millions of data possible, which has changed the way of approaching the analysis of large data sets. These changes have affected statistical analysis in diverse areas of sciences and engineering; especially domains where the spatial location plays an important role in the analysis such as geography, urban economics, public policy, environmental engineering, geology, and transportation engineering. In this section I will provide a detailed review of methodologies relevant to the scope of this dissertation.

#### 2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a specific approach of data analysis originating from the work of Tukey in the early 1970s. EDA is a preliminary exploratory examination of a data set with the aim of identifying structures or patterns within the data, checking for data quality, providing statistical summaries and plotting graphs to identify data characteristics (Chatfield, 1986). EDA helps in hypothesis generation, formulating appropriate models and provides a basis for more sophisticated analysis. The objective of EDA is to identify suitable variables and trends to formulate a model (Cox and Jones, 1981). Different types of graphs are plotted depending on the number of available variables. For univariate data sets, the stem and leaf plot, bar-charts, box-plots, histograms, dot plots, residual plots, and time series plots can be used. For bivariate



data sets, scatter-plots, Q-Q plots, level and spread plots, and mean difference plots can be used to generate useful insights; and for multivariate data - matrix and trellis scatter plots, parallel coordinate plots, 3-D scatter plots, and conditioning scatter plots can be applied. Most of these graphic techniques are useful for displaying information about data measured at the interval or ratio scale (Haining, 2009). For instance, stem-and-leaf plots display quantitative data in a graphic format, providing a quick overview of distribution. Stem-and-leaf plots retain most of the data integrity and help in detecting outliers and define distribution properties in medium sample's sizes (20-100 points). In large samples, stem-and-leaf plots become cluttered since each point is represented numerically and box plots may be a better option. Box plot displays five critical information about the data set - smallest observation, lower quartile, median, upper quartile, and the largest observation. In addition to detecting outliers, box plots can be quickly used to graphically compare distributions between data sets. More complex plots may be required for the analysis of multivariate data, as traditional descriptive statistics may only tell part of the story when dealing with multiple dimensions (Chatfield, 1986). EDA can help the analyst to identify patterns and generate insights which more rigorous classical tests cannot detect due to their stepwise nature and emphasis on specific models. However, care must be taken to ensure that the interesting patterns are not over analyzed and over interpreted (Bolker, 2008).

### **2.2.1 Robust Statistics**

Traditional statistical methods work under idealized assumptions which may not be met in practice. Robust methods include the application of resistant techniques that works under variety of conditions and not just only idealized specific conditions as is assumed in traditional

statistics. Resistant techniques provide results that are not extremely affected by outliers or departures from model assumptions. The basic concept of resistant techniques can be described as methods that provide results which are not affected significantly by the presence of atypical outlier observations (Besag, 1981). Classical methods perform poorly in the presence of outliers or when common assumptions (such as errors being normally distributed) are not met in the data. Consequently, a robust statistic, which is “resistant” to errors in the results, will produce "robust estimators" even when the assumptions are approximately met. These estimators commonly have reasonable efficiency and small bias. However they are asymptotically unbiased, and therefore the bias tends to zero as the sample size tends to infinity.

Traditionally, graphical display is the most valuable way of detecting outliers and checking if some values were erroneously recorded. However, graphical display based outlier and erroneous data detection becomes more difficult and impractical as the number of data and variables increase. Robust methods provide automatic ways of detecting, removing, and flagging outliers, putting aside the need for manual screening. Robust techniques consider the measure of data sets by values that are not greatly affected by the presence of extreme or atypical data values. It is now clearly known that the mean can perform very poorly with long-tailed distributions and that the median, among other estimators, is better with presence of long-tailed or outliers. Robust techniques use the “median” instead of the “mean” for measure of “location or center”, and the inter-quartile range instead of standard deviation as an estimator of spread distribution of numbers (Cox and Jones, 1981). The motivation is to produce estimators that are not excessively affected by small departures from model assumptions.

### **2.2.2 The Smooth and Rough Part of the Data**

A fundamental concept in the exploratory approach is that data can be partitioned into smooth and the rough component, or the fit and the residual as known in descriptive statistics. The smooth component corresponds to the predictable part of the data set values which has a consistent and regular pattern. It may be represented by a straight line or a curve describing the relationship between variables. The smooth component must be extracted from the data to obtain the rough or the unpredictable component.

$$\text{Data} = \text{Smooth} + \text{Rough}$$

Different, linear and nonlinear smoothing techniques have been developed mainly based on moving medians because they are more resistant than moving averages. For example, methods proposed by Tukey for smoothing one-dimensional series are based on moving medians of successive trios of data values, often repeated until convergence. Hanning also proposed a linear smoother with weights of 1/4, 1/2, 1/4 to smooth significant variation between variables. The objective of smoothing is to obtain a rough component with no additional pattern or structure. If the rough contains a structure not removed by the smooth, it is known as "not rough enough" and further smoothing should be conducted. This represents a difference between exploratory data analysis and classical analysis - the willingness to further smooth the rough to produce alternative models (Hartwig & Dearing, 1979).

### **2.3 Time Series Analysis**

Time series is a sequence of data values measured at specific periods of time. Time series analysis includes methods for analyzing time series data to obtain meaningful information. Time series analysis models are commonly used to predict future values based on previously observed

data. These methods of data analysis are often used to monitor industrial process or to track corporate business metrics. Time series models account for the fact that data points measured over time may have an internal structure such as autocorrelation, trend or seasonal variation that should be accounted for.

Time series analysis has been widely used in forecasting in various aspects of the transportation engineering. Williams (2003) presents a theoretical basis for modeling short-term traffic condition streams as time series formulation using the seasonal autoregressive integrated moving average processes (ARIMA). This technique was utilized for data for two representative freeways. Prediction's results by the seasonal ARIMA process were compared to three heuristic forecasting methods: the random walk forecast, the historical average forecast, and the deviation from the historical average forecast using the mean absolute percentage error (MAPE). Seasonal ARIMA was found to outperform the other heuristic methods. Pitfield (2004) examined the competitive pricing behavior of airlines, by studying whether the prices offered by one airline are correlated to the prices of other airlines or its own past prices using an ARIMA model. The objective is to determine whether prices are more closely explained by the competitor's actions or by own past price settings. Durango (2006) used time series analysis to estimate the performance of transportation infrastructure over time while accounting for the effect of maintenance, inspections and the use of technology on the life-cycle costs. Haire (2009) used time-series analysis to estimate ridership demand into the public transportation. To date there is no work on forecasting freight prices using time series analysis. Other methods such as linear and nonlinear regression analysis are primarily used. Particular studies, such as Ozkaya et. al. (2010) presents a regression methodology to estimate less-than-truckload (LTL) rates. They used an historical database of shipments in the U.S, and intangible factors such as negotiation power

in the model. Kay et. al. (2009) develops a nonlinear regression model to estimate LTL rates between origin-destination pairs in the U.S. They use public available rates as a dependent variable, and load density, shipment weight, and O–D pair distance as the explanatory variables. To date there is no work on applying time series analysis to predict truckload prices.

## **2.4 Exploratory Spatial Data Analysis (ESDA)**

Exploratory Spatial Data Analysis (ESDA) is a subset of Exploratory Data Analysis (EDA), which employs a combination of graphical and visual methods along with robust statistical techniques for exploring spatial data sets without the requirement of advance statistical knowledge. This exploration technique has the philosophy of staying close to the original data using only simple transformations and intuitive methods on the raw data without employing inference theory (Haining, 2009). ESDA helps analyst and researchers better understand the data by providing the big picture and guidance in obtaining inferences and insights. ESDA places emphasis on graphical representation of data allowing data linking and brushing, and provides a variety of views which helps the analyst determine the best method to apply. In general, the principal characteristic of ESDA is to discover spatial patterns of a data set, and should be the first and the most important step in the analysis of data (Tukey, 1977).

ESDA follows the most direct approach of allowing the data itself to reveal its underlying structure and model rather than making specific assumptions regarding the model type which best fits the data (Engineering Statistics Handbook, 2006). ESDA attempts to discover the original structure of the data set by identification of data properties, important variables, pattern's detection, anomalies, spatial autocorrelation, spatial heterogeneity, etc. that will help to formulate spatial models (Anselin, 1998).

ESDA uses linkable numerical data and geographical maps to maximize the data visualization enabling the detection of remarkable patterns that are displayed on a map (Anselin, 1999a). All these dynamical graphic methods, which began as improvements in displaying static data, e.g. scatter plots, histograms, bar charts, pie charts, etc., have allowed the user to directly manipulate of the data, which is immediately reflected into a graph or map. These graphical enhancements became likely by the availability of computer units with sufficient computational power to generate statistical graphs without delays and to allow interaction in the data through input devices, e.g. mouse, touch pen, and touch screen monitors. Furthermore, an ESDA's principle is to involve the human factor more directly in the exploration of data, exploiting the natural capabilities of the brain to detect patterns and structure and thereby gain richer insights than possible with the traditional rigid and static display. ESDA achieved these principles by allowing the user to make changes into the data set that immediately are reproduced into a map, changes such as deleting anomalies, highlighting or brushing subsections of data, establishing links between the same data points in different graphs, manipulating and projecting higher dimensional of data, all of which have the purpose of better understanding the data structure to enhance the data analysis.

Developing a model involves three stages - model formulation, model estimation, and model validation. Usually researchers expend a majority of the time and effort on the model estimation stage in estimating parameters for a given model structure. In recent times a lot of attention has been paid to the model validation stage also with respect to performing diagnostic checks, analyzing the residuals etc. However, in a lot of cases, the bigger issue is not estimating a model or validating it but identifying a model formulation. To date there is no clear procedure

for identifying the model formulation (Chatfield, 1986). ESDA can help in addressing this deficiency by providing insights for hypothesis formulation (Good, 1983).

Geographic Information System (GIS) has enabled effective visualization of geographic data as they are capable of assembling, storing, manipulating, and displaying geographically referenced information. Gahegan (1999) described some barriers into the science of visualization such as computer graphs that should have the power to deal with many layers or variables simultaneously, and the complexity of mapping between geographic data sets and the visual domain. Nowadays, these barriers are being overcome with the implementation and enhancements of GIS systems.

#### **2.4.1 Geographic Information System (GIS)**

GIS allows handling large amount of data, detecting patterns, anomalies and spatial outliers automatically. It can now perform statistical analysis, modeling, and visual display of geographic data. Thus, GIS has provided a powerful new tool that has inspired researchers to use geographic concepts and consider spatial data characteristics by explicitly recognizing the key role that distances, location, proximity, neighborhood, and region play in the society. Because of the complexity of interactions that occur in space or between spatial units, the explanation of spatial patterns and characteristics are often hard to unravel. For these reasons, new theories which include spatial effects into the statistical theory have been developed that explicitly incorporate space for dealing with these complexities and for facilitating inferential judgments (Anselin et al. 2004a).

## **2.5 Data Visualization**

Visualization of the data sets is divided into three types of graphical representation: data mapping, flow maps, and autocorrelation maps.

### **2.5.1 Data Mapping**

Data mapping can help identify unique characteristics in Truckload prices. For instance, there may be substantial difference in TL prices for the same delivery distance between pair of states at different locations. This difference in prices can be potentially due to cross-subsidization or price discrimination in the TL market at certain locations. Plotting data into a map helps to identify these characteristics, and answer important questions as described below.

- Which states had been affected by high, low, and medium TL prices, considering the origin and destination states?
- What states are subjected to large or short delivery distance?

### **2.5.2 Representing Interaction between Origin-Destination Trips**

It is a fact that geographical movements have a tremendous positive/negative impact at locations where they have occurred. For instance, the movement of people, money, commodities, could increase traffic congestion, improve business, decrease quality of life etc. Representation of geographical movement in statistical models is a complex process but can help in interpreting significant characteristics across space. Understanding origin-destination interactions help in discovering travel behavior, providing knowledge of major and minor locations under supply and demand, which ultimately, can be aggregated for modeling purposes. Recently, graphical and



visualization maps have been developed which help in exploration of data interaction between origin and destination.

In the truckload market, interaction of deliveries can be graphically represented by logistic lanes, from origin to destination U.S. states. A graphical representation of the origin destination interactions can help in identifying the surplus and deficit areas. Surplus areas are considered key production zones, usually, at these locations most of deliveries begin; deficit areas are locations where production is insufficient or does not meet local demand for specific commodities. Identifying these locations play an important role in economic studies, suggesting the attractiveness and unattractiveness of those locations for economic activities e.g. labor, migration, etc. Generally, freight movement of commodities across space are from surplus to deficit areas. It is important to point out the various constraints that roadway freight movement deals with across geographic space and periods of time over a year, e.g. climate, congestion, roadway closure, etc.

In general, flows are represented by objects such as lines and arrows, where thickness represents the quantity of the flow, and arrows indicate the direction of the flow. In large data sets, these objects have negatively impacted the visualization and interpretation of maps, because they become clutter, and thus, it makes them unable for distinguishing important locations and patterns. Currently, alternative visual objects have been investigated to reduce density into the graph, improving the human perception.

Tobler (1987) developed the pioneering software in this area on a Windows-based platform. It has an updated version, Flow Mapper 1.1, released in 2004 for The Center of Spatially Integrated Social Science (CSISS). Recent computer packages in this area are: Flowmap (2010) at the Utrecht University; FlowMapping (2010) for The Spatial Data Mining

and Visual Analytics Lab at the University of South Carolina; Jflowmap (2012) at the University of Fribourg. They allow users for visualizing and exploring data interactions, as well as, for performing data analysis. Jflowmap was selected in this research due to its large data handling capabilities, offering various visualization approaches for producing and analyzing flow maps. It effectively represents flows through good contrasted colors, which allows an outstanding human perception. The requirement for using Jflowmap is custom programming; the development of GIS shapes files and data preparation.

Jflowmap represents flows by lines, instead of arrows, lines' edges are red and green colored to represent the origin and destination respectively. The thickness indicates the volume of the flow, which in this research is the number of deliveries between locations. The largest number of trips are drawn above smaller flow volumes. Jflowmap allows users to customize data filtering to obtain detail visualization and summarization of relevant data. Also, it allows for highlighting and merging high spot locations (nodes) for aggregating and displaying purposes. Jflowmap uses a flow bundling technique to reduce visual clutter and make maps more understandable. The algorithm of bundling is based on attracting and putting together nearby flows with similar positions and orientations. Thus, flows turn into joint paths similar to electrical wires (Boyandin, 2010). This technique improves the users' visualization, facilitating revealing significant patterns such as high and low consumer and demand locations, and also for identifying busy regions connected by trade. The use of multiple maps is advisable when working with large datasets recorded at different periods of time (Boyandin, 2011).

This research develops flow maps for truckload movements for each year, from 2005 to 2010, and for the total number of movements during the six year period. Flow maps help in responding important questions as described below.

- What are the major sources and destinations of the flows?
- Where is the largest and the smallest flow?
- In which direction and states are major movements?
- What is happening within a specific area?

### **2.5.3 Spatial Autocorrelation**

The first law of geography states “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). This law highlights the concept of autocorrelation across space in data analysis. ESDA helps to measure the degree or significance in which near and distant observations are related. Based on statistical test and graphical representation, similarities among observation can be identified across locations. These similarities can be positive, negative or have zero relationship. Positive relationship occurs when similar values happen in nearby locations, negative relationships when dissimilar values happen in nearby locations, and zero when the relationship is by random chance. A prior knowledge of a spatial weight matrix (SWM) is required for this analysis. SWM creates a spatial relationship (weight) between observations, establishing which neighboring unit will be averaged. Selecting an appropriate SWM is an important and complex task for the spatial analysis formulation. It can be driven by a rational analysis, prior knowledge, or exploring the data sets to obtain insight regarding what type of connectivities fit best the data set. It is important to mention that the selection of an appropriate SWM is mostly related to modeling perspective, because the idea is captured both, the spatial and non-spatial effects of the dependent variable (GeoDa, 2012). ESDA can attempt to provide insights related to which SWM would be appropriate; therefore, for modeling purpose, it is advisable a sensitivity analysis of various SWM specifications

(Anselin, 2005). ESDA helps in identifying similarities and dissimilarities of the dependent variable at various locations. GIS tools allow for identifying and visualizing spatial autocorrelation into a map. GEODA software developed by Dr. Luc Anselin is used in this research for this purpose. It helps in displaying the dependent variable across locations, providing tools for univariate and bivariate statistical analysis, and allowing users for brushing and linking charts and maps.

## **2.6 Spatial Econometric Analysis**

Spatial econometric is a sub field of econometric with its origins in the early 1970s in Europe. The origin of spatial econometric theory recognizes that data collected at different locations may not be independent as usually is suggested in conventional statistics. These techniques were developed due to the need of taking into account spatial autocorrelation and spatial heterogeneity in regression models for cross-sectional and panel data (Paelinck and Klaassen, 1979; Anselin, 1988a). Initially, research that incorporated geographic factors in the models were primarily focused on specialized areas such as economic geography, regional sciences, urban and real estate economics Anselin (1999b). However, over the last decade spatial econometric methods have been applied to a wide variety of domains. Two major factors motivate the need to specify, estimate and test the presence of spatial interaction between variables: (i) the need for a method to deal with spatial interactions and (ii) the recognition of failure of standard econometric techniques when dealing with the presence of spatial autocorrelation (Anselin, 1999b).

The large progress in geographic information systems (GIS) and the availability of geo-data in many disciplines have created the need of specialized methods to deal with geographic

data. The use of GIS with data collected at various locations, and spatial modeling techniques is a mutual place for various applied sciences in engineering, economics and policy analysis. Anselin (1999b) describes four areas of interest in spatial econometrics: (i) the correct incorporation of spatial effects in econometric models; (ii) the estimation of these models; (iii) the development of tests to determine the presence of spatial effects; and (iv) spatial forecasting or prediction. This study will be focused on the last three concerns, since they fall in the research questions of the study.

Various techniques are found in the literature to deal with spatial dependency and to specify spatial models. However, the two most prominent approaches are the spatial autoregressive model (SAR), in which the dependent variable is affected by the geographic location of the others dependent variables, and the spatial error model (SEM), in which the error term is affected by error correlations across space. In other words, autocorrelation can be modeled by considering correlation among the dependent variable (SAR) or by considering correlation among the error terms (SEM). Extensions of those models are developed by including higher order of dependency, and combined dependency across the dependent variable, independent variables, and residuals. For instance, the Spatial Durbin Model (SDM) is an extension of a SAR model, which considers spatial dependency in both, the dependent and independent variables. Spatial models incorporate a spatial weight matrix (SWM) to account for spatial dependency across neighborhood zones. The SWM attempts to capture the hidden effects that geographical location has over neighboring zones. The most common SWM encountered in the literature are: the contiguity matrix, the k-nearest neighbor's matrix, a connectivity matrix, and a distance based matrix (LeSage and Pace, 2009). Other studies specify SWM in order to capture the origin and destination effects or both, for instance in interaction or flow models

(LeSage et al., 2009). Usually SWM is standardized to have a row sums equal to one; this is required to produce linear combinations of effects from neighboring regions of the model.

During the last decades the analysis of time series and cross-sectional data has been widely used in quantitative studies of comparative political economy (Podesta, 2000). This combination of cross-sectional data at specific interval of times such as having multiple snapshots was referred to pooled analysis (Pennings et al. 1999) or now commonly named panel data models. Panel models deal with various sizes of data in terms of periods of time and number of units. They also consider testing data by two dimensions, considering all units throughout time.

In similar fashion, the spatial econometric literature includes a two dimensional analysis, considering dependency similar to the SAR and SEM models. These models are called spatial panel models, which are being developed and empirically tested by various authors such as Debarcy (2011), Elhorst (2011), Baltagi (2011), Yu (2011), LeSage et al. (2010), and Anselin et al. (2005). The main objective is developed a model which is able to examine simultaneously the space-time dimension.

Spatial econometric techniques have not been extensively applied in the freight transportation domain. LeSage and Pace (2005) provided a methodology to model origin-destination commodity flows incorporating spatial dependence into the traditional gravity model used in international trade. LeSage and Polasek (2006) introduced a regression-based gravity model for commodity flows, incorporating information regarding the highway network by introducing a spatial weight connectivity matrix. It includes information regarding the presence or absence of a major highway/train corridor that passes through the regions. Results indicated that this approach improves the model fit, providing higher likelihood values than using typical

spatial weight matrices, such as the described by LeSage and Pace (2005). Behrens et al. (2007) proposed a new version of the gravity equation, a dual model with cross sectional interdependence and spatially lagged error terms. An empirical application using trade data between Canada, and the U.S. is provided. Results indicated that controlling directly for spatial interdependence across trade flows, significantly reduces border effects, since it captures multilateral resistance. Novak et al. (2008) investigated freight generation models, using commodity data at the national level. By comparison of methods, results indicate that the spatial approach is a superior technique than the linear freight approach in order to specify freight generation models. This study concludes that the existence of spatial issues needs to be recognized and addressed in modeling efforts. Cho (2005) examines patterns of rural land development and density using spatial econometric models and Geographical Information System (GIS). The study infers that this type of spatial analysis can capture broader physical and social phenomena than traditional least squared analyses, which may miss. Results revealed that closer distance to roads and cities, greater access to streams and rivers, higher elevations, and greater proportions of a flat area are valued highly in rural land development. Yong (2006) applied spatial econometric to study the effect of investing in highway infrastructure has over the economy of Missouri. The study investigated the effect on retail trade, wages, employment growth in manufacturing, and in attracting people. Results indicated that highway investment does not provide significant positive evidence into attracts business or establish firms in the region. However, it has a favorable effect to increase wages rather than to increase labor supply. Quddus (2008) developed a series of relationships between many area-wide traffic casualties, and the contributing factors associated with ward characteristics in the Greater London metropolitan area. Both, non-spatial models and spatial models were developed to identify the

similarities and the differences among these relationships. Results suggested that since crash data are collected concerning location measured as points in space, spatial dependence exists between the area-level crash observations; therefore, spatial models are more appropriated for modeling those relationships. Safett (2009) conducted a global and local spatial regression analysis to evaluate the safety performance of roads in the provinces of Turkey. Road accident rates and death rates from 2001 to 2006 were used as dependent variables; and number of motor vehicles and length of roads, among others, were used as independent variables. Results indicated high concentration of fatal accidents related to roads connecting three biggest cities in Turkey. The predicted values of accident and death rates by spatial regression analysis showed better results than ordinary least regressions.

Previous empirical studies using spatial econometric have concluded that since data in the transportation field is collected at various spatial locations, spatial econometric performs better than conventional statistics. Dealing with the spatial characteristics of a data set significantly helps in obtaining better results, and in capturing the spatial effects across locations. The disadvantage of this method in the transportation field may be centered on specified the appropriate SWM. Practitioners can define various types of SWM for dealing with spatial data, but results may not be unique and may vary accordingly with the SWM specification. Increase efforts to define an adequate SWM may be required depending on the type of study.

## **2.7 Spillover Impacts**

Spatial spillover refers to a situation in which the alteration of one unit affects neighboring units. This is an indirect effect associated with an activity considered as an externality, which can be positive or negative to the neighbors. Neighbors can be defined by



geographic characteristics as distance or location, and also by social network. An important issue of analyzing spatial spillover is to identify the diffusion channels, how the externalities spreads across locations, which is considered a black box in the literature (Capello, 2009). Krugman (1991) states that “knowledge flows are invisible; they leave no paper trail by which they may be measured and tracked”. For instance, high unemployment rates tend to reduce consuming demand in a region, available income of consumers is affected; therefore, the purchase of goods and services is also reduced. It slows the economy of the region and consequently impact into neighboring regions. The last statement illustrates the propagation of effects across locations due to the alteration of one variable in a specific location. Spatial models have made possible of quantifying the spatial spillover impacts across locations. LeSage and Pace (2009) provides a methodology to measure these direct and indirect (spillover) impacts. They propose an N-by-N cross-partial derivative matrix; where the off-diagonal elements represent the indirect impact due to the change in the  $r^{th}$  explanatory variable in one region impacts the dependent variable in neighboring region. The diagonal elements represent the direct effects. They also propose a scalar summary measure for both effects. It is important to remark that this method is associated with spatial lag models, spatial correlation in the dependent variable, e.g. SAR or SDM. For SEM models the spillovers impacts are directly measured by the coefficient estimates in the same fashion as ordinary least squares.

Studies that consider spatial econometric techniques for measuring the indirect impacts across neighboring locations have been conducted, primarily in the regional economic field. Rey (1999), in his US regional income converge study, investigated the impact of how a shock in one state replicates on surrounding states, complicating the convergence process. Baumont (2001) studied the spillovers effects of growth in European regions. This study indicates that the overall

growth of a region is positively influenced by neighboring regions, and by simulation showed the impact of a shock in one region disperses to all the regions of the sample. Crandall (2004) studied the spillover effects of poverty, results indicate that poverty is tied to wealth neighboring areas; thus, reducing poverty in particular areas will reduce poverty in neighboring locations.

In the transportation field, there are few studies in the literature. Parent and LeSage (2010) quantify the spatial spillover effects in terms of commuter travel time when infrastructure capacity of a specific roadway segment is increased within the roadway network. Yang (2012) investigated the spillover impacts in inbound and domestic tourism flows. The study confirmed both impacts, and evokes that infrastructure and tourist attractions have a significant spillover impacts between cities in different regions. Zenhua (2012) quantifies the impact of surface transportation infrastructure such as highways, public railways and public transit in a region. This study confirms the significant impact of infrastructure to the economy of a region, mainly reached by the spillover effects from these facilities. Hu et al. (2010) quantifies the indirect effects of investing in transportation infrastructure in one region contributed to the GDP of neighboring regions, results indicate a significant contribution to the economic growth.

## **2.8 Forecasting with Spatial Econometric**

In the transportation engineering field, forecasting is an important part for designing and planning existing and future transportation facilities. The forecasting of movements of goods and people as well as prices take an important role in the design stage considerations. Little attention has been paid in defining appropriate forecasting methods, because of the assumption that state departments and local agencies are aware of the transportation needs for the future (Wells, 2009). The forecasting techniques vary from state to state and by local agencies, and usually these

standard statistics provide unsound results. Consequently, forecasting skills are becoming an important issue for federal transportation agencies. “Agencies will no longer be able to be agnostic about what forecasting techniques work and which do not” (Wells, 2009). For instance, prices in freight transportation are considered fixed for specific routes, and associated with factors such as fuel cost and delivery distance. Standard models consider them, but miss in considering economical and geographical factors such as the demand and supply associated with the transport activity and the indirect effects affecting other locations.

Forecasting using spatial panel models is not extensive in the spatial econometric literature. In the transportation field, no available studies have been found. Empirical and experimental studies have been conducted. Baltagi (2011) used Monte Carlo simulation to compare sixteen forecast panel data models and found that that ignoring spatial effect lead to inaccurate forecast outcomes. Miller (1997) investigated the accuracy of regional employment forecasts using spatial panel data models and found the forecast values to be very accurate at the state level with MAPE values less than 2.0 percent; where MAPE is the mean absolute percentage error, which is used to measure accuracy in terms of percentage, the lower the percentage values represent higher accuracy. Mayor (2012) evaluated short-run forecast of unemployment in Switzerland and Spain using spatial panel data models and obtained MAPE values less than 6.0 percent.

## CHAPTER 3

### 3.0 METHODOLOGY

#### 3.1 Introduction

This research applies exploratory, statistical, and econometric techniques to answer the research questions. The analysis starts with the exploratory analysis. EDA and ESDA techniques will allow us to understand the data sets, providing insights on significant patterns and variations from past years in the truckload price market and help in determining correlations over space and time. The data set is checked for errors, missing values, and observations which are not consistent with the rest of the data. Exploratory analysis helps in providing insights regarding potential assumptions in data sets, such as normality and constant variance in a data set are at least not unreasonable assumptions. Spatial autocorrelation for the dependent variable TL price (price-per-mile) is tested by the Moran's I and LISA test. Scatter plots, matrix plots and maps help to visualize and understand patterns in freight truckload price per mile data. At the final stage of this process, descriptive summary statistics and graphs are supplied which provides useful insights regarding the relationships between the dependent and independent variables. Statistical methods such as time series analysis and robust techniques from EDA are applied to identify significant patterns and to estimate price-per-mile truckload. A comparison chart between methods is provided based on accuracy. Patterns discovered through ESDA approach are the starting point, providing insights with respect to clustering the states into regions. It helps us to formulate spatial econometric models that allow us to estimate TL prices at the state level and to quantify spillover impacts across states.

## 3.2 Time-Series Analysis

Truckload prices have large variations over time and are affected by other variables such as diesel cost and distance, which are noise or unpredictable variables. Diesel cost shows high fluctuations over months and years. Time series analysis can be used to analyze high noise data sets with data points obtained at fixed points over time because they can capture the underlying structure of the data providing a path to estimate future values. Changes and variations over time can be described as trends or cycles. In economics, it is conventional to decompose time series into a variety of components, some or all of which may be present in a particular instance as shown below.

$$Y = T + S + u$$

where:  $Y$  are the observations over a specific time period,  $T$  is the slowly changing component or global trend;  $S$  is the periodic component or the seasonal variation; and  $u$  is the residual or the disturbance term. When the previous components, the trend and the seasonal are extracted from the observations, the noise or residual is obtained, which is the rough or irregular part for which no single explanation can be provided (Pollock, 1993). The residual may have some form of autocorrelation. Therefore, having some knowledge about the residuals in some observations can help to formulate better models and thus increase forecasting accuracy.

### 3.2.1 Autoregressive and Moving Average Process

Autoregressive models and moving average are often used in the analysis of univariate time series data. AR(p) models follow the process shown in equation 1.

$$X_t = \delta + \theta_1 \cdot X_{t-1} + \theta_2 \cdot X_{t-2} + \dots + \theta_n \cdot X_{t-p} + u_t \quad (1)$$

where  $X_t$  represents the dependent variable collected at specific period of time,  $u_t$  is the disturbance term,  $\theta_i$  are the parameters in the model, and  $\delta$  represents the constant term that captures the effects not obtained by the explanatory variables. In general, it is a linear regression considering as explanatory variables past dependent variables, which can be the most recently past observation or prior values of the series. The value of  $p$ : is called the order of the AR model. In the other hand, the moving average (MA) models follow the process shown in equation 2.

$$X_t = M + u_t + \theta_1 \cdot u_{t-1} + \theta_2 \cdot u_{t-2} + \dots + \theta_p \cdot u_{t-p} \quad (2)$$

where  $X_t$  represents the dependent variable, which follow a time series process,  $M$  is the mean of the series,  $u_{t-i}$ ,  $i = 0 \dots p$ , are the disturbances and  $\theta_i$  are the parameters in the model. The term  $p$  is named the order of the MA model. MA follows a linear regression, but dissimilar to the AR models, the explanatory variables are the disturbance from past observations of the series. Disturbances are considered normally distributed with zero mean and constant variance.

### 3.2.2 Autocorrelation Data Diagnostic

Diagnostic tests for autocorrelation can provide useful information about the "order" or the number of lags involved in the dynamic process of the data. The autocorrelation can be in the error terms or in the dependent variables. Understanding the degree of autocorrelation in the data helps in formulating the correct time series model. Equation 3 shows a simple first order autoregressive process (AR1) in error terms.

$$Y_t = \beta \cdot X_t + e_t ; e_t = \rho \cdot e_{t-1} + v_t \quad (3)$$

In the above equation is important to note that  $v_t$  is independent and identically distributed. Equation 4 shows a simple first order moving average (MA1) in error terms.

$$Y_t = \beta \cdot X_t + e_t ; e_t = \rho \cdot v_{t-1} + v_t \quad (4)$$

Autocorrelation plots such as autocorrelation function (ACF) and partial autocorrelation (PACF) plots are developed by graphing the dependent variable. They help to examine how the dependent variable propagated over time, allowing for identifying the numbers of AR and/or MA terms or lags that explain better the dependent variable. For the time series analysis, the dependent variable is the truckload price per mile and the explanatory variables considered are diesel cost and delivery distance. The model is calibrated using data from a five-year time period from 2005 to 2009. The most recent year data of 2010 is used to validate the model.

### **3.2.3 Forecasting Values Procedure**

The dependent variable (price-per-mile) is averaged for similar months over the five-year period. The residuals are obtained by extracting the trend and the seasonal component. The trend is obtained by ordinary least squares (OLS), by regressing the dependent variable over time. The trend provides a baseline for the value of the dependent variable, depicting the general way in which the price per mile fluctuates over the course of a year. The detrend values are obtained by subtracting the trend from the actual cost per mile for each month. The seasonal component captures the seasonal variation in the price per mile for each month. The seasonal values are the average of the *detrend* values, calculated for similar months, e.g. January 2005, January 2006... over the five-year period, 2005-2009. These seasonal factors are constant in the model for each month. Having the trend and seasonal factors, the residuals at each month can be obtained by subtracting those values from the dependent variable. After obtaining the residuals of each observation, the mean and variance can be calculated; they can be used to simulate future residuals.

The simulation process is conducted assuming autocorrelation of the error term,  $u_t = \rho.u_{t-1} + e_t$  (error-lag) as shown in Equation 2. The  $\rho$  parameter and the standard deviation of the error ( $s$ ) is estimated by regressing the residuals at time  $t$  to those at  $t - 1$  over the five-year period. The error term ( $e_t$ ) is simulated from the normal distribution of zero mean and calculated standard deviation. Therefore, by Equation 2, and using the mean and standard deviation of the residuals, will allow for forecasting residuals for future months. Finally, the model is completed after obtained the trend, the seasonal factors and the underlying structure of the residuals over time. MATLAB will be used to develop a routine to forecast residuals and to estimate the models for each of each datasets.

### 3.3 Smoothing Techniques

Every time series  $\{Y_t: t = 0,1,2..,n\}$  can be split into a smooth structure component  $\{S_t: t = 0,1,2..,n\}$  and a rough unstructured component  $\{R_t: t = 0,1,2..,n\}$ . Examples of smoothing are the linear smoothers (equation 5), Hanning smoothers (equation 6) and median smoothers (equation 7).

$$S_t = \frac{Y_{t-1} + Y_t + Y_{t+1}}{3} \quad (5)$$

$$S_t = \frac{1}{4}Y_{t-1} + \frac{1}{2}Y_t + \frac{1}{4}Y_{t+1} \quad (6)$$

$$S_t = \text{median} (Y_{t-1}, Y_t, Y_{t+1}) \quad (7)$$

Note that the median smoothing can be conducted using 3 or 5 consecutive terms. Sometimes smoothing is applied multiple times. Applying one smoother to the results of a previous smoother is known as re-smoothing. A useful re-smoothing procedure is denoted as 3R where 3 denotes that median smoothing using 3 consecutive terms is applied and R denotes that the smoothing is repeated until no further changes are noted.



Sometimes running median smoothers remove interesting patterns. In such cases, the rough obtained by residual from the re-smoothing process is smoothed and added to the smoothed sequence. Often the same smoother is used in both smoothing, and re-roughing. When this occurs, the smoother has been used “twice”, for example 3R twice. Flat segments may generate by two points; those will be smoothed by splitting (S). The idea is to replace each of the two points of the segment by a median calculated by considering the point itself and points from the right and left side.

In this study two smoothing techniques will be considered, the 4253 H twice, and the 3RSS H twice. The 4253H twice consists of running medians of 4, then 2, then 5, then 3 followed by Hanning smoothing. The result of this smoothing is then re-roughed by computing residuals, applying the same smoother to them and adding the result to the smooth of the first pass. The 3RSSH, twice consisting of running median of 3, two splitting operations named S to improve the smooth sequence, each of which is followed by a running median of 3, and finally, Hanning. The end points are dealt with using the method described by Velleman and Hoaglin (1981). The full smoother 3RSSH, twice is produced by re-roughing as described above.

### **3.4 Model Validation and Accuracy**

The withheld-data validation method is used in this study. The most recent available data for the year 2010 is used for validating purpose. The accuracy of the model will be obtained by comparing the forecast values and actual price-per-mile from the validation year. Mean Absolute Percent Error (MAPE) will be used to measure accuracy. This measure is generally used for time-series analysis and allows for comparing different models. MAPE is defined as:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{(Y^i - Y_f^i)}{Y^i} \quad (8)$$

where  $Y^i$  is the actual value from the dataset and  $Y_f^i$  is the forecasted value. Commonly, in the industry and business application, a model is considered to provide highly accurate forecasting if MAPE values are smaller than 10%, and the model is considered to provide good forecasting if the MAPE values are between 10% and 20% (Lewis, 1982).

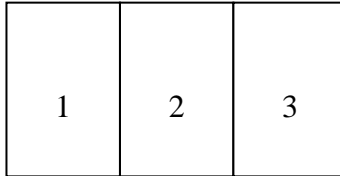
### **3.5 Exploratory Spatial Data Analysis (ESDA)**

ESDA approach is performed using the average of price-per-mile for each state and year. TL price data range from 2005 to 2010, and missing values exist for some states at each data set as described in the analysis. The objective of this analysis is to discover spatial patterns that help in identifying spatial clusters and formulating accurate forecasting models. The investigation regarding spatial autocorrelation in the data is also conducted as a part of ESDA; thus, a spatial weight matrix is defined for this purpose.

#### **3.5.1 Spatial Weight Matrix**

To evaluate spatial autocorrelation, the closeness of observations must be defined based on different forms of vicinity such as near neighbors, contiguity, distance, etc. Normally the vicinity between various observations is captured using a spatial weight matrix (SWM) which defines the relationships between locations where measurements were made. For instance, if data is collected at  $n$  locations, then the spatial weight matrix will be of dimension  $n \times n$  with zeros on the diagonal. The weight matrix can be specified in many forms such as constant weight for any two different locations, fixed weight for all observations within a specified distance,  $k$ -nearest neighbors with a fixed weight, and all others zero, proportional weight to distance or inverse distance between zones. Variations of weight matrices such as squared distances or log-

distance are also possible. The spatial weight matrix is row-standardized - the sum of the weights of a row is equal to one. The following is an illustration of a contiguity based weight matrix for 3 regions as shown in figure 1.



**FIGURE 1: Region's Locations**

The contiguity spatial weight matrix, indicate whether any pair of observations are neighbors, neighbors are defined when two regions share common borders. Therefore, if region  $i$  and region  $j$  are neighbors, then,  $w_{ij} = 1$  and zero otherwise. Using this illustration above, and the first-order contiguity definition, locations which share a border, the SWM have zeros on the main diagonal, avoiding of define an observation as neighbor to itself, and on position associated with non-contiguity. Ones in the matrix indicate contiguity neighbors. In order to normalize the influence across locations, the matrix often is row-standardized, in which the sum of each row is equal to one, providing proportional weights at each location. Figure 2 illustrates the first-order contiguity spatial weight matrix.

$$W = \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{matrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \xrightarrow{\text{row standardized}} W = \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{2} \\ \mathbf{3} \end{matrix} \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

**FIGURE 2: Row-Standardized First-Order Contiguity Matrix**

### 3.5.2 Measures of Spatial Autocorrelation

Moran's I (Moran 1950) tests for global spatial autocorrelation for continuous data. This test considers the null hypothesis: "observations are randomly distributed among locations" or in other words, the spatial process is a random chance; thus, a zero value indicates a random process. Equation 9 illustrates the cross-products of the deviations from the mean, which is calculated for  $n$  observations for a variable  $X$  collected at various locations (indexed by  $i$  and  $j$ ) as:

$$I = \frac{n \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{S_0 \sum_i (X_i - \bar{X})^2} \quad (9)$$

where  $\bar{X}$  is the mean of the variable  $X$ ,  $w_{ij}$  are the elements of the weight matrix, and  $S_0$  is the sum of the elements of the weight matrix:  $S_0 = \sum_i \sum_j w_{ij}$ . Moran's I varies from -1 to +1. In the absence of autocorrelation the expectation of Moran's I statistic is  $-1/(n-1)$ , which tends to zero as the sample size increases, regardless of the specification of the weight matrix. For a row-standardized spatial weight matrix, the normalizing factor  $S_0$  equals  $n$  (since each row sums to 1), and the statistic simplifies to a ratio of a spatial cross product to a variance. A Moran's I coefficient larger than  $-1/(n-1)$  indicates positive spatial autocorrelation, and a Moran's I less than  $-1/(n-1)$  indicates negative spatial autocorrelation. Thus, Moran's I of -1 and +1, indicates perfect dispersion and perfect correlation respectively, a zero value indicates random spatial pattern. Moran's I is sensitive to extreme values, being a preferred measure of spatial autocorrelation. Cliff and Ord (1981) have shown that Moran's I is consistently more powerful than other similar tests. This index is enhance when is used as Local Indicator of Spatial Autocorrelation (LISA) proposed by Anselin (1995). LISA considers breaking down Moran's

Index and assigns a LISA index to each location; thus, the sum of LISA's for all observations is proportional to the Moran's I.

LISA helps to interpret local spatial patterns, and to check for anomalies in the global spatial pattern. Moran's I test and LISA are visualized on scatter plots and maps, respectively. These graphs provide a visual perception of local and overall patterns. The scatter plot is developed using the spatial lag on the vertical axis, and observations at each location in the horizontal axis. The slope of the regression line advocates the existence of global autocorrelation, and local autocorrelation is advocated by examining local trends in the scatter plot. The LISA map helps to observe clustering areas. This scatter plot has four quadrants; thus, observations are classified in four categories. Categories related to TL prices can be described as: High price states with High price neighbors (HH); Low price state surrounded by High price neighbors (LH); High price state with Low price neighbors (HL); and Low price states with Low price neighbors (LL). HH and LL correspond to positive forms of spatial dependence; HL and LH represent negative spatial dependence.

### **3.6 Spatial Econometric Models**

The spatial models are used for modeling cross-sectional data samples, considering the integration of effects due to the geographical location of the data. LeSage (1997) and Pace et al. (1998), LeSage (1999) provides a comprehensive treatment of these models from a maximum likelihood perspective. The most general statement of a spatial autoregressive model is shown in equation 10.

$$y = \rho W_1 y + X\beta + u \quad (10)$$

$$u = \lambda W_2 u + \epsilon$$

$$\epsilon = N(0, \sigma^2 I_n)$$

where  $y$  contains a  $n \times 1$  vector of cross-sectional dependent variables,  $X$  represents a  $n \times k$  matrix of explanatory variables and  $W_1$  and  $W_2$  are  $n \times n$  spatial weight matrices, usually containing contiguity relations or functions of distance as previously described. Based on the general model, special models can be derived by imposing restrictions. Setting  $X = 0$  and  $W_2 = 0$  produces a first-order spatial autoregressive model in the form of:

$$y = \rho W_1 y + \epsilon \quad (11)$$

$$\epsilon = N(0, \sigma^2 I_n)$$

This model attempts to explain variation in  $y$  as a linear combination of contiguous or neighboring units with no other explanatory variables. It represents a spatial analogy to the first order autoregressive model from time-series analysis,  $y_t = \rho y_{t-1} + \epsilon_t$ ; where total reliance is on past period observations to explain variation in  $y_t$ . Setting  $W_2 = 0$  produces a mixed regressive-spatial autoregressive model. This model is analogous to the lagged dependent variable model in time series. Equation 9 with additional explanatory variables ( $X$ ) to explain variation in  $y$  over the spatial sample of observations takes the form of equation 12.

$$y = \rho W_1 y + X\beta + \epsilon \quad (12)$$

$$\epsilon = N(0, \sigma^2 I_n)$$

Letting  $W_1 = 0$  results in a regression model with spatial autocorrelation in the disturbances shown in equation 13.

$$y = X\beta + u \quad (13)$$

$$u = \lambda W_2 u + \epsilon$$

$$\epsilon = N(0, \sigma^2 I_n)$$

A related model is known as the spatial Durbin model where a “spatial lag” of the dependent variable as well as a spatial lag of the explanatory variables matrix  $X$  is added to a traditional least-squares model.

$$y = \rho W_1 y + X\beta_1 + W_1 X\beta_2 + \epsilon \quad (14)$$

$$\epsilon = N(0, \sigma^2 I_n)$$

As previously described this research considers the spatial dependence in the data; therefore, spatial econometric analysis is conducted for modeling TL prices.

### 3.6.1 Spatial Panel Models

Panel data refer to the analysis of data sets that combine time series with cross-section dimensions; thus, the analysis considers the  $N$  cross-sections, e.g. states, regions, firms; and the  $T$  number of time periods. Panel data has the important characteristic that provides coefficient estimates with lower standard errors than a single cross-section data, because of the large number of observations throughout time. It allows for controlling each specific individual over time, and for controlling unobserved heterogeneity, which could generate bias coefficient estimates when using a standard ordinary least square (Soderbom, 2011). Panel data is categorized based on the number of cross sections and time periods in the data set. For instance, a data set with the same time periods for all individuals is called a balanced panel; and a data set with different lengths of the time period across individuals is called unbalanced panel, the last are the most usual type of data sets. A panel data with missing observation (unbalanced panel) can become a balanced panel by rejecting the missing observations on the dataset; however, it may decrease the efficiency and accuracy of the coefficients estimates. An important issue in modeling panel data

is how the unobserved individuals are considered in the model. They can be correlated with the regressors (fixed effects), or they can be distributed independently of the regressors (random effects). Equation 15 illustrates both concepts.

$$y_{it} = \alpha_{it} + X_{it}\beta + \varepsilon \quad (15)$$

where,  $y_{it}$  is the vector with the dependent variables for each cross section at the specific period of time  $t$ .  $X_{it}$  contains the explanatory variables ordered by cross-section and periods of time,  $\alpha_{it}$  is the unobservable variable (effect), and  $\varepsilon$  is the error term.

Fixed effect assumes that  $\alpha_{it}$  is correlated to  $X_{it}$  and it is time-invariant. Hence by subtracting equations  $\alpha_{it}$  is eliminated, allowing consistent estimates of  $\beta$ . The random effect assumes that  $\alpha_{it}$  is iid $[0, \sigma^2]$  and hence is uncorrelated with  $X_{it}$ . The random assumption is stronger to the fixed effect, which provides consistent coefficient estimates. However, if the true model involves fixed effects, the random effects will be inconsistent (Cameron and Trivedi, 2005, 697). A standard test that helps identifying between fixed and random effects is the Hausman test, which is described in the next section.

The spatial panel model considers assumptions similar to the standard panel model, but adds the spatial dependency in either the dependent variable or the residuals comparable to the SAR and SEM models. Similarly, the Durbin model specification includes the constructed variable  $WX$ , which represents spatial dependency in the explanatory variables. Equation 16 and 17 displays the general form of the spatial panel lagged Durbin model and spatial error Durbin models respectively.

$$y_{it} = \alpha_{it} + X_{it}\beta + WX_{it}\beta_d + \rho W y_{it} + e; \quad (16)$$

$$e \sim N(0, \sigma^2 I_n)$$



In the above equations,  $y_{it}$  represents the dependent variables ordered by cross-section and by time period,  $X_{it}$  contains the  $k$  explanatory variables in the model, and  $\beta$  is the coefficient estimate for each explanatory variable. The parameter  $\alpha_{iN}$  represents either the random or fixed intercept in the spatial panel model,  $\rho$  is the spatial autoregressive parameter, and  $W$  is the SWM.

$$y_{it} = \alpha_{iN} + X_{it}\beta + WX_{it}\beta_d + u; \quad (17)$$

$$u = \rho Wu + \varepsilon;$$

$$\varepsilon \sim N(0, \sigma^2 In)$$

The spatial Durbin error model omitted spatial dependency in the dependent variable, but considers the spatial dependency in the residuals.  $u$  is the residual,  $\varepsilon$  is the disturbance terms, the description of the other parameters and coefficients are similar to equation 16.

### **3.7 Spatial Test for Spatial Dependency and Effects**

Literature regarding spatial panel models identified two major tests that help analysts in specifying a model. The Lagrange Multiplier (LM) test examines dependency either in the dependent variable or in the residuals. The Hausman test helps in determining the most suitable effect in the model between fixed and random effects.

#### **3.7.1 The Lagrange Multiplier Test**

Anselin (1988b) developed the Lagrange Multiplier (LM) diagnostic test for spatial dependence. LM test evaluates for spatial misspecification in the data in either the dependent variable or in the disturbance. It also tests for spatial heterogeneity, the existence of uneven distribution across space. The disadvantage of LM is that only one particular type of misspecification is tested, which may lead to erroneous information if the true model includes

both the presence of spatial lag and spatial autocorrelation in the residuals. Bera and Yoon (1993) developed a robust LM test under local miss-specification. This test basically corrects the mean and covariance matrix of the standard LM statistic, being simpler to compute. Anselin, Florax and Yoon (1996) introduce a similar test for spatial models, having the advantage of simple computations, and being suitable to identify spatial dependence in the response variable and in the disturbance. Anselin, Gallo and Jayet (2008) developed similar LM tests for spatial panel models. This research conduct the standard and robust LM tests to identify spatial-lag and spatial-error model formulations. In the case that both the LM-lag and LM-error test are significant, the model with the larger LM value (t-statistics) is selected, because it entails a better and more reasonable model (Anselin, 2005). LM test follows the  $\chi^2$  distribution with one degree of freedom and it is asymptotic in nature. The null hypothesis of the LM-lag test states that no significant presence of spatial autocorrelation exists in the dependent variable as is shown in equation 18.

$$LM_{lag} = \left( \frac{e'Wy}{\sigma^2} \right)^2 / B \quad (18)$$

where  $B$  takes the form of equation 19.

$$B = \left[ (WXb)' \cdot \frac{(I-X(X'X)^{-1}X')(WXb)}{\sigma^2} \right] + tr(W^2 + W'W) \quad (19)$$

$b$  is the OLS coefficient estimates (betas);  $X$  is the  $n \times k$  matrix with explanatory variables;  $e$  is the residuals from the OLS;  $W$  is the spatial weight matrix,  $\sigma^2$  is the OLS variance. The null hypothesis of the LM-error test states that no significant presence of spatial autocorrelation exists in the residuals as is shown in equation 20.

$$LM_{error} = \left( \frac{e'We}{\sigma^2} \right)^2 / tr(W^2 + W'W) \quad (20)$$

where  $e$  constitutes the OLS residuals;  $W$  is the SWM of  $n \times n$  dimension, and  $tr$  is the matrix trace operator (sum of the elements of the main diagonal). The robust test as described before adjusted the mean and covariance matrix of the standard LM statistic.

### 3.7.2 Random vs. Fixed Effects (Hausman Test)

The null hypothesis of the Hausman test states that the effects ( $\alpha_{it}$ ) are distributed independently from the explanatory variables ( $X_{it}$ ), a random effect process. The main objective is to detect violation of the random effect assumption. Therefore, if no correlation exist between the explanatory variables and the effects, then the coefficient estimates in the fixed effects model ( $\hat{\beta}_{FE}$ ) and in the random effects model ( $\hat{\beta}_{RE}$ ) should be similar. Equation 21 shows the null hypothesis ( $H_0$ ), a measure of the differences between the two coefficient estimates.

$$H_0 = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' \cdot [Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})]^{-1} \cdot (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \quad (21)$$

$H_0$  is chi-square distributed with degrees of freedom equal to the number of regressors in the model. Therefore, if the null hypothesis is rejected (p-value<0.05), rejection of the random effects, indicates that the estimates for the two models are significantly dissimilar, in favor of the fixed effects model. The random effect will be biased and their coefficients estimates inconsistent and significantly different when compare to the fixed effects estimates (Dougherty, 2007, pp417). Contrarily, accepting the null hypothesis (p-value>0.05) indicates rejection of fixed effects; thus, the coefficient estimates in the random and fixed effects are considerably similar. In this case, both the fixed and random effects are consistent, but fixed effects will be inefficient, because it includes unnecessary dummy variables that drop the degrees of freedom. However, it does not necessarily imply that these effects are the true effect in the model, free from bias, and therefore preferred over the fixed effects estimator. The Hausman test can

potentially fail due to the lack of statistical power to consistently determine departures from the null hypothesis (Clark, 2012). Empirical studies have shown the unlikeliness of obtaining true zero correlation between exploratory variables and effects; thus, random effects will still be biased even if the Hausman test is in favor of the null hypothesis. Clark (2012) indicates that a biased estimator can provide a greater reduction of variance than an unbiased estimator; therefore, the former could be preferred. Analysis using spatial panel models requires working with data collected at different spatial locations, usually contiguous units across continuous periods of time, to provide with efficient and consistent results and to define a SWM. Elhorst (2011) indicates that under this situation, the fixed effect is the preferred effect in the model. He states that “the idea that a limited set of regions is sampled from a larger population must be rejected and therefore the random effects models”. In the spatial econometric literature, researchers usually report both effects attempting to compare coefficient estimates.

### **3.8 Spatial Spillover Impacts**

The main contribution of the spatial econometric theory is that makes possible the measurement of spatial spillover impacts. It can quantify the impact of changing one explanatory variable in the model, attributed to a specific location or period of time, to the dependent variable in neighboring locations. Conventional linear regression models do not allow for measuring impacts across units. LeSage and Pace (2009) documented the theory behind the estimation of the direct and indirect effects for SAR models. It is important to mention that estimation of indirect impacts is exclusively for lagged models; SEM models do not allow this measure. The direct and indirect effects are measured through the matrix of partial derivative ( $M$ ) of the dependent variable  $Y$  with respect to the  $k$  explanatory variable at specific  $n$  locations. The direct

effect is the average of the diagonal elements of  $M$ , and indirect effect is the differences between the direct effects and total effects, the latter is the average of either the sum of rows or sum of columns of the non-diagonal elements of  $M$ . In other words, the sum of the diagonal elements is the direct impacts due to changes in the  $k^{th}$  explanatory variable; and the sum of rows of the off-diagonal elements is the indirect impact affecting neighboring locations. Equation 23 represents the matrix  $M$ . The valuable of this measure is that provides two scalar numbers for the direct and indirect effects across locations (LeSage and Pace, 2009). Interpreting and reporting results are less complicated with these numbers, since displaying the total  $M$  matrix of  $n$ -by- $n$  elements is not required.

$$y = \alpha \cdot i_n + \rho \cdot W \cdot y + X \cdot \beta + \varepsilon \quad (22)$$

$$\partial y / \partial x_k = (I_n - \rho \cdot W)^{-1} \cdot I_n \cdot \beta_k \quad (23)$$

Similarly, for the Durbin model the direct and indirect impacts come out from equation 25.

$$y = \alpha \cdot i_n + \rho \cdot W \cdot y + X \cdot \beta + W \cdot X \cdot \theta + \varepsilon \quad (24)$$

$$\partial y / \partial x_k = (I_n - \rho \cdot W)^{-1} \cdot (I_n \cdot \beta_k + W \cdot \theta_r) \quad (25)$$

LeSage and Pace (2009) described the advantage of the partial derivatives in Durbin models, being less restrictive by allowing deeper impacts across space than SAR models.

## CHAPTER 4

### 4.0 EXPLORATORY DATA ANALYSIS (EDA)

#### 4.1 Introduction

Truckload price data consists of an original database developed by a third party logistic (3PL) company. The data set is classified into four categories according to the type of goods and the lane logistics management: (i) Dry-managed, goods are dry and lanes are managed (3PL manages the truckload logistics operations) (ii) Dry non-managed – goods are dry but other 3PL company is managing the truckload logistics operations, (iii) Reefer managed – non-dry goods and lanes managed, (iv) Reefer non-managed – non-dry goods and other 3PL company is managing the truckload logistics operations. The data sets include truckload price-per-mile from origins to destinations, average distance and the date recorded from 48 States in the U.S., between 2005 and 2010. In this research monthly average of truckload price-per-mile is used to conduct a time series based analysis. In addition, the impact of fuel price (diesel price) is considered; diesel price was obtained from the Energy Information Administration (EIA) for each state, month, and year (<http://www.eia.gov>). Table 1 illustrates a data sample.

**TABLE 1: Data Sample**

Obs.#	From	To	Year	Month	Day	Distance	Price/Mile	Fuel_Ori	Fuel_Dest
1	AL	FL	2005	1	1	404	1.89	1.801	1.821
2	WI	CO	2005	4	1	1224	1.55	2.171	2.212
3	IL	CO	2005	4	1	987	1.45	2.171	2.212
4	IL	CA	2005	4	1	2081	1.5	2.171	2.368
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1543	AL	CA	2008	1	31	1997	1.2	2.316	2.274
1544	OH	GA	2008	1	31	658	1.6	2.279	2.351
1545	GA	IL	2008	2	8	630	0.98	2.333	2.235
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35564	GA	IL	2009	2	13	724.8	0.98	2.333	2.235
35565	NY	FL	2009	2	20	1159	1.799	2.426	2.333
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
43750	TX	OR	2010	2	24	2192	1.43	2.231	2.340
43751	KS	TX	2010	2	27	790	1.5	2.235	2.231

**4.2 Descriptive Statistics**

Table 2 illustrates descriptive statistics for each data set over the five year period.

**TABLE 2: Dataset Descriptive Statistics**

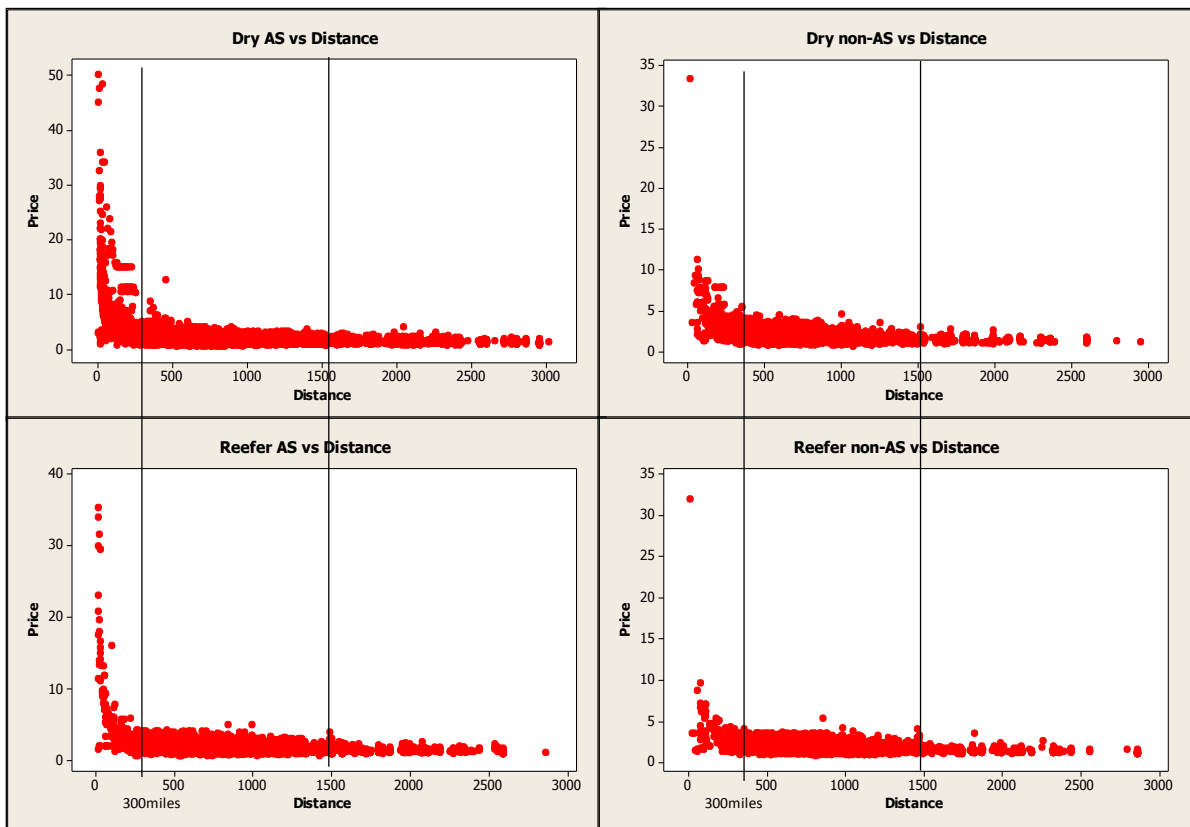
Type of Load	Variable	Mean	SE Mean	St Dev	Coef. Var	Minimum	Median	Maximum
<b>Dry Managed</b>	Av. Cost	1.557	0.0295	0.2288	14.69	1.185	1.5066	2.3154
	Av. Dist.	895.5	35.2	272.6	30.44	370.8	857.6	1725.4
	Fuel Ori.	2.854	0.0835	0.647	22.67	1.949	2.7047	4.7091
	Fuel Dest	2.889	0.0838	0.649	22.46	2.019	2.7562	4.8391
<b>Dry non-Managed</b>	Av. Cost	1.559	0.0265	0.2051	13.15	1.1875	1.5507	2.206
	Av. Dist.	929.7	25.9	200.6	21.57	552	896.3	1506.2
	Fuel Ori.	2.861	0.0836	0.6479	22.64	1.9347	2.7296	4.744
	Fuel Dest	2.889	0.0825	0.6389	22.12	1.9981	2.7412	4.725
<b>Reefer Managed</b>	Av. Cost	1.523	0.021	0.1624	10.66	1.1767	1.4831	1.9294
	Av. Dist.	1282.7	35.9	278	21.67	657.5	1242.3	1974.8
	Fuel Ori.	2.861	0.0831	0.6437	22.49	1.9745	2.7057	4.7042
	Fuel Dest	2.897	0.082	0.6355	21.94	2.0255	2.7501	4.7592
<b>Reefer non-Managed</b>	Av. Cost	1.560	0.0174	0.1351	8.66	1.2742	1.5685	2.0299
	Av. Dist.	1205	28.5	220.9	18.33	671.7	1167.7	1863.7
	Fuel Ori.	2.862	0.0833	0.6452	22.54	1.9792	2.7029	4.716
	Fuel Dest	2.896	0.0825	0.6388	22.05	2.0367	2.7123	4.7903

Table 2 indicates that for the four types of data over the five year period, from 2005 to 2009, the average price-per-mile is around \$1.55 per mile. The average delivery distance varies between 900 for dry loads and 1,300 miles for reefer loads. Higher variability is observed in price-per-mile of dry goods compared to reefer goods. Managed lanes data have higher variability in price-per-mile when compared to the non-managed lanes data. Exploratory analysis of the raw data was conducted for each data set as shown in the next section.



### 4.3 EDA Plots

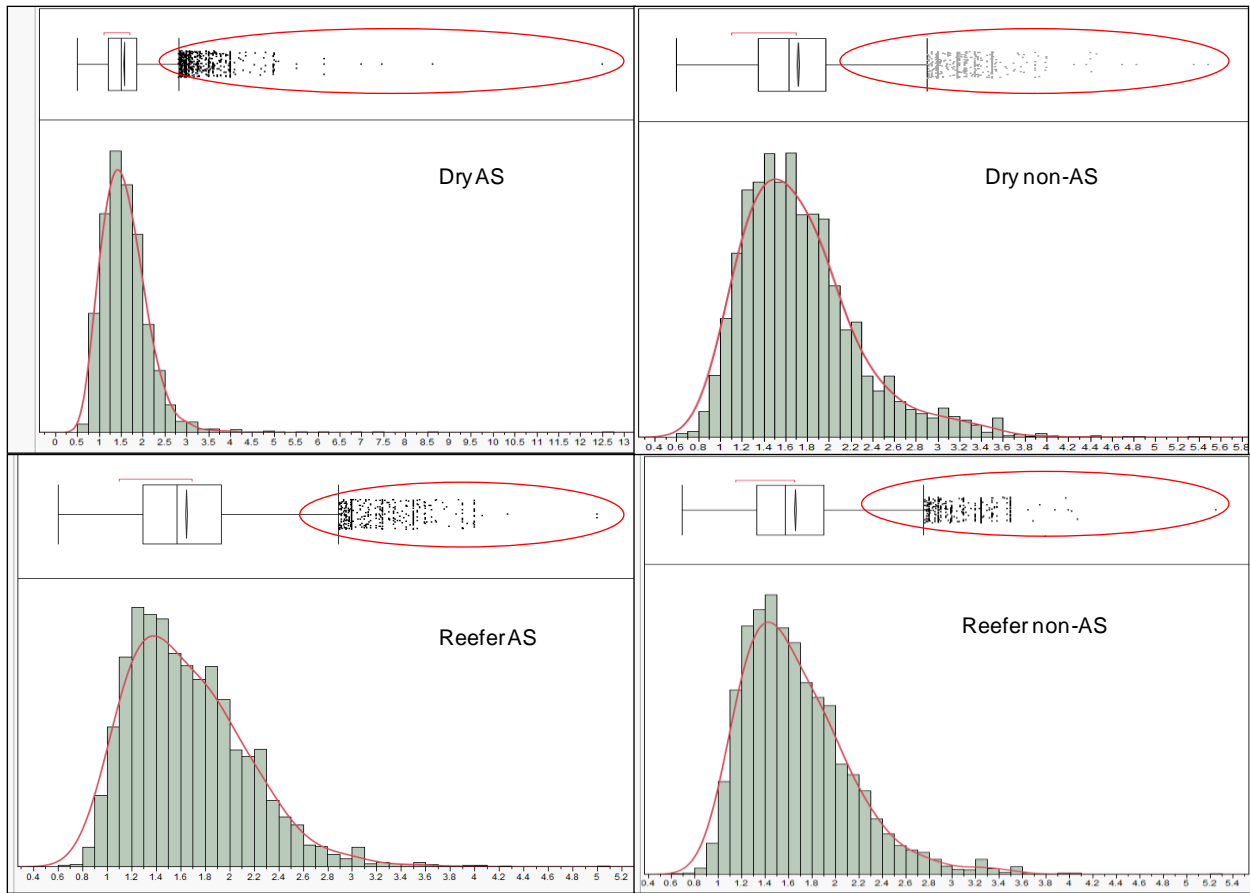
Exploratory analysis of the raw data was conducted for each data set. A scatter plot of price vs. distance is shown in figure 3. The original four data sets consist of real price-per mile for 115,201 origin-destination pairs. Dry managed accounts for 38.56 percent, Dry non-managed for 14.18 percent, Reefer managed for 32.64 percent, and Reefer non-managed for 14.63 percent of the total number of observations.



**FIGURE 3: Scatter Plot- Price vs. Distance**

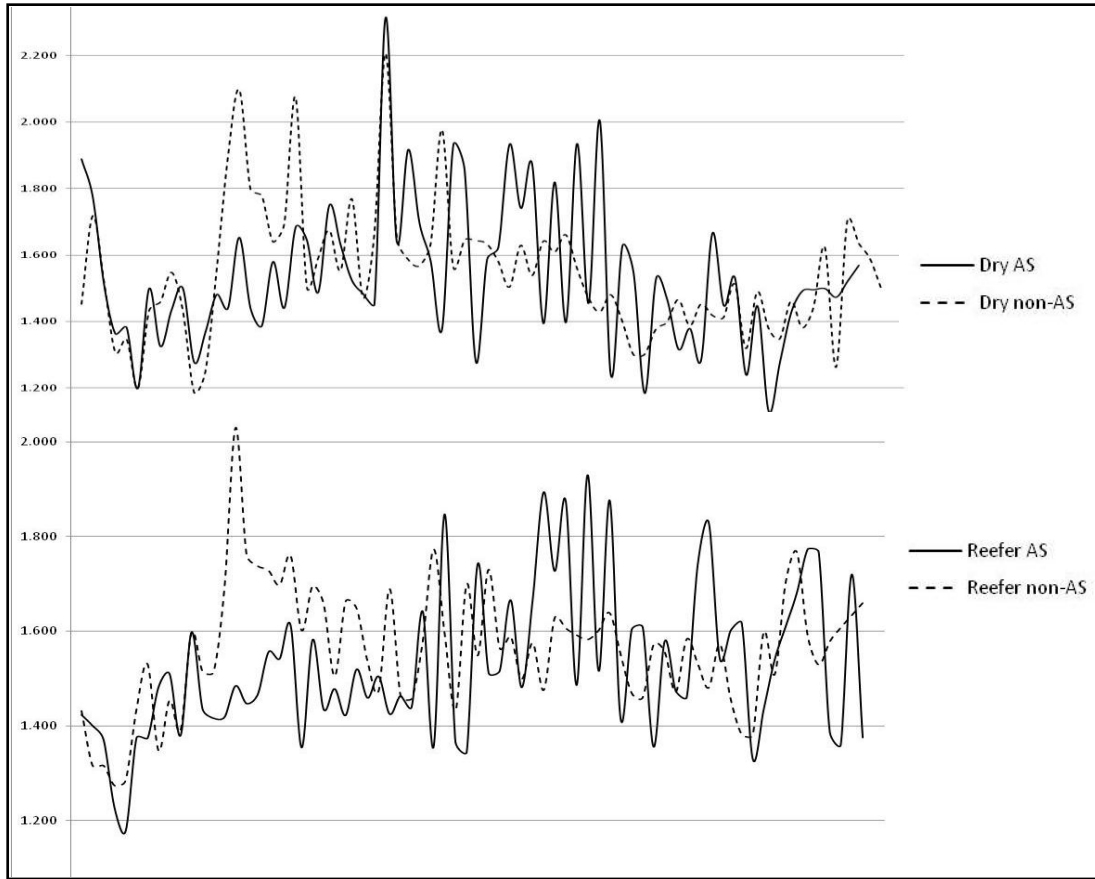
Figure 3 shows the relationship between price-per-mile and distance. It is observed that prices for distance smaller than 300 miles have large variation. TL prices are between \$10.00 and

\$30.00 per mile in delivery distances lower than 300-mile. This is because truckload prices over short distances vary significantly depending on local factors and supply chain practices. Normally a flat rate is charged for shorter distances based on the above mentioned factors. This section is focus on developing models for delivery distances greater than 300 miles. After removing distance greater than 300 miles, the data sets consist of 107,322 origin-destination pair observations. Dry managed accounts for 37.06 percent, Dry non-managed for the 12.20 percent, Reefer managed for the 35.03 percent, and Reefer non-managed for the 15.70 percent of the total number of observations. Based on the above scatter plot, two time series models were developed for forecasting truckload prices. The first model (termed model A from now on) was for predicting truckload prices for distances greater than 300 miles. In addition, it is observed that most of the deliveries were made between the 300 and 1500 miles. Thus, a specific model for that range was developed (termed model B from now on), which is expected to provide higher forecasting accuracies. Note that because of the lack of data, a model associated with deliveries distances greater than 1500 miles can not be performed. The details of the time series models is provided in chapter 5. A distribution diagram and box plots of prices related to distances greater than 300miles is shown in figure 4.



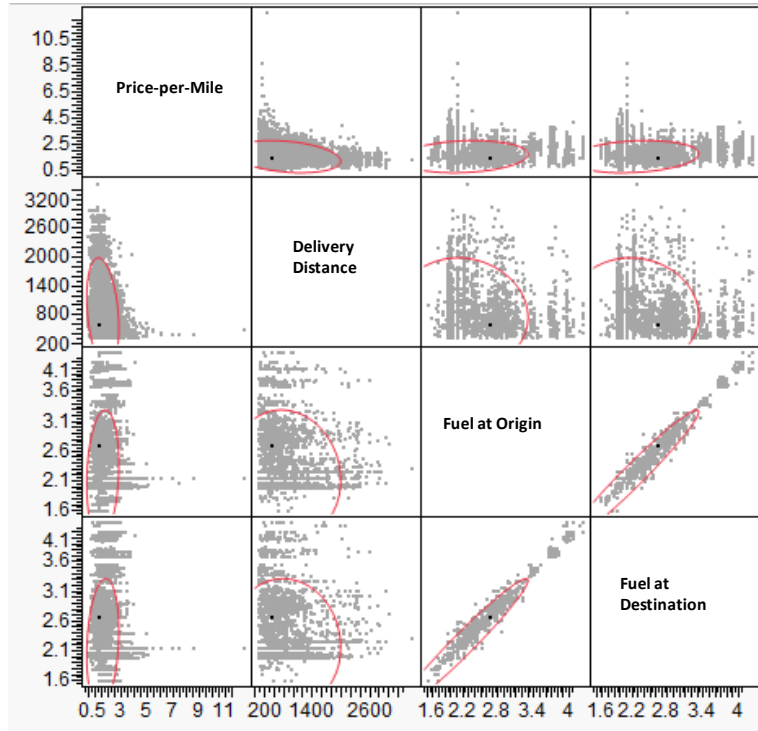
**FIGURE 4: Distribution Diagram and Boxplot of Price-per-Mile**

Figure 4 indicates positively long tail distribution (skewed) for all data sets. It displays the existence of extremely large or small observations in the data sets, which are affecting the mean. It suggests that robust techniques, which are resistant to outliers, could be an appropriate approach for modeling purposes. A monthly price-per-mile for each dataset over time is shown in Figure 5.



**FIGURE 5: Price Fluctuation over Time**

Figure 5 indicates a large variations with different maximums and minimums at same periods of time for each dataset which validates the need for a specific model for each dataset. Figure 5 also shows a lot of random fluctuations over time which suggests the applicability of a time series based analysis. A multivariate scatter plot of Dry managed lanes with price-per-mile, fuel cost at origin, fuel cost at destination and delivery distance is shown in figure 6.



**FIGURE 6: Multivariate Scatter Plot for Dry Managed**

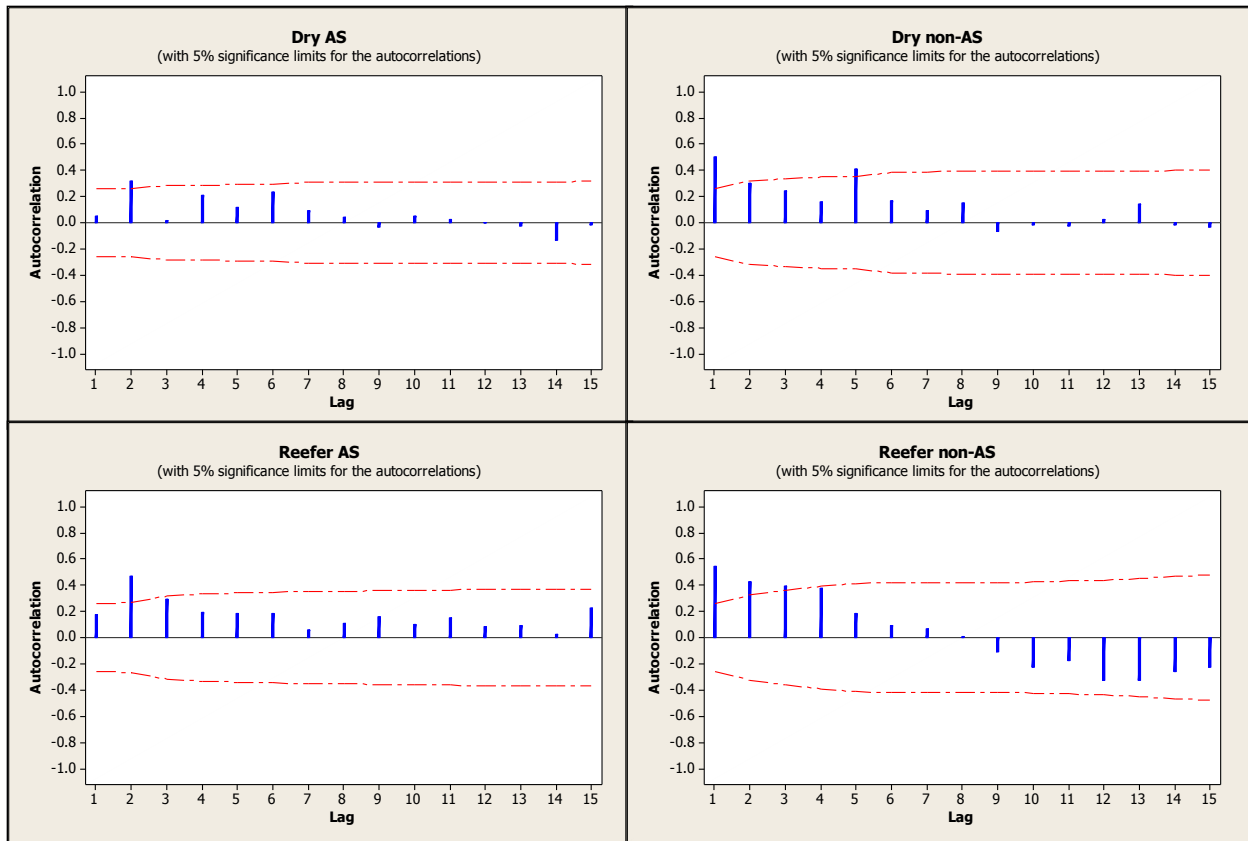
Figure 6 indicates a linear relationship between fuel cost at origin and destination . It suggests the use of only one variable among fuel cost at origin and fuel cost at destination for modeling purposes to avoid collinearity problems. It is also observed that price has a random relationship with other variables, showing no specific pattern. Similar insights were obtained for the three data sets. It is important to mention that fuel cost at origin and at destination were tested in the time series models, and both have shown similar results. Therefore, only the model with fuel cost at origin is reported. The matrix plot also indicates that the response variable (price-per-mile) does not have a linear relationship with other variables, displaying a random fluctuation. It suggests that considering data in the form of time series could be a good approach for modeling the data set. Figure 7 illustrates a steam-and-leaf plot for the dry managed data.



The scatter plot indicates that there is a relationship between distance and price-per-mile. Small price-per-mile is most likely for large distances. The brushing and linking tool indicates that prices for distance smaller than 300 miles are above \$5.00 /mile. The scatter plot also indicates that the prices between \$3.00 and \$5.00 are spread out up to a maximum distance of 1500 miles. It is as well observed that prices for distance smaller than 300 miles have large variation. The box plot and scatter plot of "distance" indicates that for distance greater than 1500 miles the prices seem to be steadier. In addition, it is observed that most of the deliveries were made between 300 and 1500 miles, thus, a specific model for that distance range could be led to provide high-accuracy estimate values. Therefore, the analysis suggests that data can be split into three groups, for distances between 0 and 300 miles, and distances between 300 and 1500, and distance larger than 1500 miles.

#### **4.4 Data Diagnostic Plots**

Diagnostic plots were developed for each data set. Autocorrelation function (ACF) and partial autocorrelation (PACF) plots are developed by graphing the truckload price-per-mile dependent variable. The plots help in identifying the numbers of lags that better explain the dependent variable. Figure 9 illustrates the ACF plots.



**FIGURE 9: Autocorrelation Plots**

Figure 9 displays the autocorrelation plots for the different data sets. The plots resemble an AR process as in almost all cases the ACF decreases as the lag increases. The dry and reefer non-managed appear to be under the lag (1) process, while the dry and reefer managed seems to be under no lag and lag (1) or lag (2) process respectively. In this dissertation, the AR(1) process was chosen, since it provides an adequate and consistent approximation of the real behavior of the dependent variable over the time period.



## 4.5 Discussion and Conclusion

Results indicate an average price-per-mile around \$1.55 per mile, and average delivery distance of 1,500 miles for each data set. Dry goods and managed lanes data have higher variability than reefer goods and non-managed lanes data, respectively. Delivery distances smaller than 300 miles have a large variability in prices, prices are above \$5.00 per mile and most of the observations fall in between \$10.00 and \$30.00 per mile. Short delivery distances have higher prices, these deliveries produce less vehicle-mile-traveled, which increase cost of factors such as in vehicle storage. Thus, on short distance carriers usually have flat rates which depend on local conditions. A positive long tail was observed for each data set, suggesting a log transformation for normalizing the data for modeling purpose. It also displays the presence of extremely large or small observations in the data sets affecting the mean; thus, robust techniques resistant to outliers seem an appropriate approach for modeling purposes. Monthly data have shown a large variation with different maximums and minimums at the same periods of time for each dataset, it validates the need of a specific model for each dataset. The matrix plot indicates a non-linear relationship between the price-per-mile and other variables. The fuel cost either from the origin or the destination of the deliveries is not significant in the models. A random fluctuation over time between these data sets suggests the time series based analysis as an appropriated approach for modeling the data. In addition, the autocorrelation plots indicates the AR (1) process as the most adequate and consistent approximation of the real behavior of the dependent variable over the time period.

Lanes with larger delivery distance were found to have lower TL prices. Plots displays prices between \$3.00 and \$5.00 for delivery distances smaller than 1500 miles. Most of the data fall within this range. At delivery distances greater than 1,500 miles prices look steadier;

however, the data points corresponding to this distance range is less. Results suggest the development of two models, model A for distances greater than 300 miles, and model B for deliveries distances between the 300 and 1500 miles. A model for distance greater than 1500 miles is not feasible, because of the lack of data associated with these deliveries distances.

## CHAPTER 5

### 5.0 TIME SERIES ANALYSIS

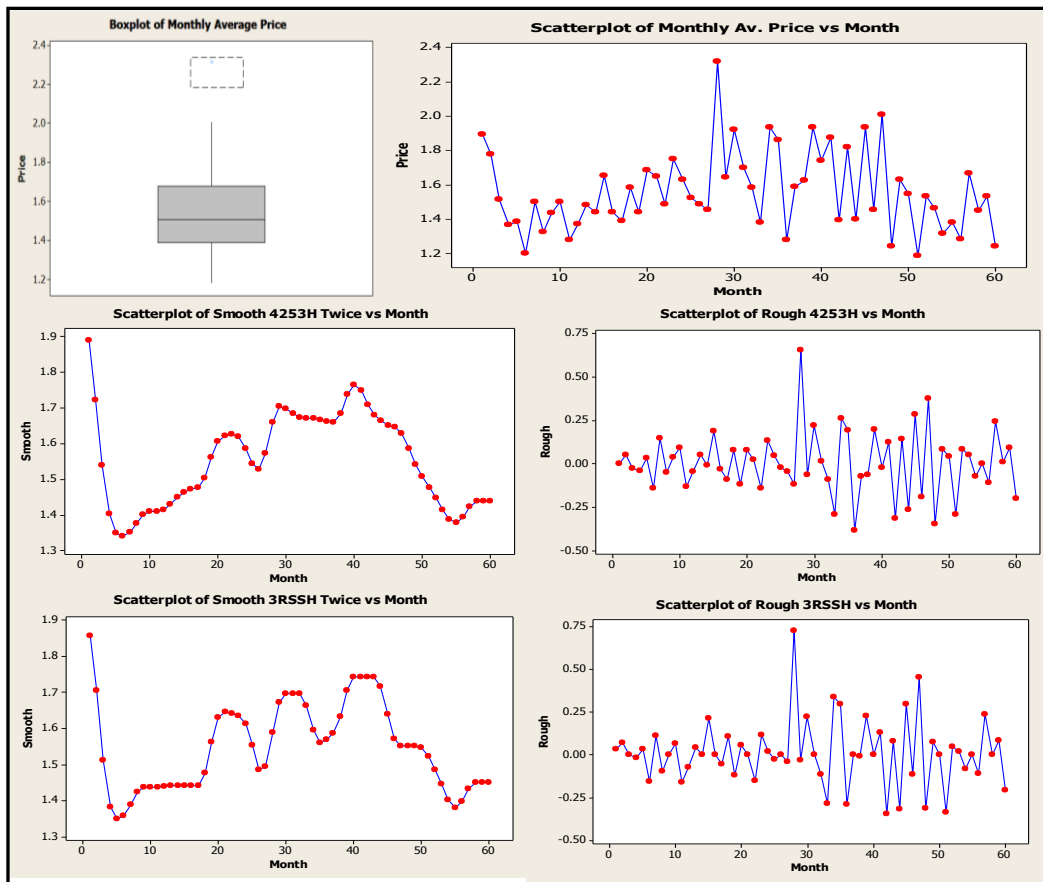
The data used in this section is the same from the EDA analysis. Data were averaged on a monthly basis to conduct a time-series analysis. Data is separated into four categories: Dry managed lanes, Dry non-managed lanes, Reefer managed lanes, and Reefer non-managed lanes. In addition, the analysis has included the fuel price data (diesel price) obtained at monthly basis from the Energy Information Administration (EIA) for each state, month and year (<http://www.eia.gov>). This section investigates multiple methods for forecasting truckload price-per-mile in the United States. The models developed were calibrated and validated on a real world truckload price-per-mile dataset. Two models were developed depending on the delivery distances as described in the previous chapter: (i) Model A which focuses on delivery distances greater than 300 miles and (ii) Model B which is calibrated and validated on delivery distances greater than 300 miles and lesser than 1500 miles.

#### 5.1 Autoregressive Model (AR), Model Calibration and Estimation

First order autoregressive process AR (1), was calibrated for each dataset over the five years time period, from 2005 to 2009. The data for 2010, which is the most recent data available, was withheld to validate the model. The methodology process is described in section 3. The next section provides an overview of the robust techniques used for predicting truckload price-per-miles.

## 5.2 Smoothing Process

In addition to the autoregressive process, a smoothing process as described in the methodology section was conducted. Figure 10 illustrates the 4253 H twice and 3RSSH twice smoothing techniques of the TL price data over 5 years (60 months). The main idea of this process is attempted to visualize any pattern in the smooth and the rough part of the data. Two smoothing techniques were conducted, the 4253H twice and the 3RSSH twice.



**FIGURE 10: Smoothing Techniques**

Figure 10 indicates that the 4253H seems to reduce more the jagged variations than the 3RSSH twice smoothing technique, which present some flat segments. The 4253H follows best

the data points, smoothing the pick values, and at the end month (60) seems to have a best fit than the 3RSSH. The rough plots indicate that the 4253H has a noise that does not follow any interest pattern. The 4253H twice looks as the best approach among the techniques presented in the analysis.

### **5.3 Results**

In the forecasting process, similar to the time series, in the robust techniques, the data were separated based on distances. The first 60 months of data is used for calibrating the model and the last 12 months for validating the model. The data is first smoothed using two techniques the 4253H twice and the 3RSSH twice to obtain the smooth part. The autocorrelation plots of the smoothed time series were also found to show the presence of an AR (1) process. Therefore, the time series analysis techniques described in the previous section were applied to the smoothed data. Table 3 shows the calibration and validation performance for model A which contains data for all delivery distances greater than 300 miles for the following six model specifications. The first three model specifications use time index (months) to calculate the trends whereas the last three model specifications use time index (months) and the fuel cost (\$/gallons) to calculate the trends. Within each model specification there are three types of models: (i) AR(1) time series model, (ii) AR(1) time series models applied on data smoothed using 4253H twice smoothing techniques, (iii) AR(1) time series models applied on data smoothed using 3RSSH twice smoothing techniques. Table 4 shows the same results for Model B which contains data for delivery distances between 300 and 1500 miles. The accuracy of the model was obtained by comparing the forecast values and actual price-per-mile from the validation year using the Mean Absolute Percent Error (MAPE).

**TABLE 3: Model A Results Comparison**

<b>Variables</b>		<b>Time (Months)</b>			<b>Time (Months) + Fuel Cost (\$/gal)</b>		
<b>Technique</b>		<b>Time Series</b>	<b>Robust</b>	<b>Robust</b>	<b>Time Series</b>	<b>Robust</b>	<b>Robust</b>
		<b>AR-1</b>	<b>4253H</b>	<b>3RSSH</b>	<b>AR-1</b>	<b>4253H</b>	<b>3RSSH</b>
<b>Dry</b>	Cal (MSE)	0.0081	0.0329	0.0354	0.0088	0.0336	0.036
<b>Managed</b>	Val (MSE)	0.0396	0.0381	0.036	0.0374	0.0357	0.0323
	Val MAPE)	<b>11.09%</b>	<b>10.75%</b>	<b>10.84%</b>	<b>10.73%</b>	<b>9.42%</b>	<b>9.24%</b>
<b>Dry non-Managed</b>	Cal (MSE)	0.0038	0.0185	0.0174	0.004	0.0187	0.0176
	Val (MSE)	0.0277	0.0209	0.021	0.028	0.0216	0.0225
	Val (MAPE)	<b>9.84%</b>	<b>8.57%</b>	<b>8.29%</b>	<b>10.20%</b>	<b>8.15%</b>	<b>8.07%</b>
<b>Reefer Managed</b>	Cal (MSE)	0.0038	0.0151	0.017	0.0036	0.0149	0.0168
	Val (MSE)	0.0453	0.0383	0.029	0.0405	0.0347	0.028
	Val (MAPE)	<b>11.09%</b>	<b>9.90%</b>	<b>9.19%</b>	<b>10.83%</b>	<b>9.75%</b>	<b>9.21%</b>
<b>Reefer non-Managed</b>	Cal (MSE)	0.0005539	0.006	0.0065	0.00069	0.0062	0.0066
	Val (MSE)	0.0137	0.0135	0.0119	0.0147	0.0138	0.0118
	Val (MAPE)	<b>5.94%</b>	<b>5.48%</b>	<b>5.32%</b>	<b>6.04%</b>	<b>5.53%</b>	<b>5.20%</b>

**TABLE 4: Model B Results Comparison**

Variables		Time (Months)			Time (Months) + Fuel Cost (\$/gal)		
		Time Series	Robust	Robust	Time Series	Robust	Robust
Technique		AR-1	4253H	3RSSH	AR-1	4253H	3RSSH
<b>Dry</b>	Cal (MSE)	0.0075	0.031	0.033	0.008	0.0315	0.0334
	AS	Val (MSE)	0.0347	0.0333	0.0313	0.0308	0.031
	Val (MAPE)	<b>9.73%</b>	<b>9.41%</b>	<b>9.14%</b>	<b>9.07%</b>	<b>8.34%</b>	<b>8.15%</b>
<b>Dry</b>	Cal. (MSE)	0.0042	0.0191	0.0195	0.0042	0.0191	0.0195
	non-AS	Val (MSE)	0.0203	0.0144	0.0161	0.0219	0.0166
	Val (MAPE)	<b>8.08%</b>	<b>6.31%</b>	<b>6.55%</b>	<b>8.34%</b>	<b>6.43%</b>	<b>6.73%</b>
<b>Reefer</b>	Cal (MSE)	0.0036	0.0157	0.016	0.0032	0.0153	0.0156
	AS	Val (MSE)	0.044	0.0386	0.0384	0.0407	0.0339
	Val (MAPE)	<b>10.40%</b>	<b>10.43%</b>	<b>10.18%</b>	<b>9.96%</b>	<b>9.53%</b>	<b>9.57%</b>
<b>Reefer</b>	Cal (MSE)	0.00064	0.0079	0.0081	0.000721	0.008	0.0082
	non-AS	Val (MSE)	0.0116	0.011	0.0104	0.0121	0.0122
	Val (MAPE)	<b>5.37%</b>	<b>5.45%</b>	<b>5.28%</b>	<b>5.53%</b>	<b>5.90%</b>	<b>5.70%</b>

#### 5.4 Performance Evaluation of Methodologies

Both the regular time series analysis and the combination of robust smoothing techniques with time series analysis provide satisfactory results in forecasting truckload prices. The application of robust smoothing techniques before performing the time series analysis was found to yield better forecasting accuracies. The MAPE using robust techniques were found to be less than 10% in all the cases when fuel cost was used as an explanatory variable in addition to the time index. When the only time index was used the MAPE for robust techniques was found to be less than 10% in all the cases except for Reefer managed data set. However, no significant improvements in MAPE were obtained using the extra explanatory variable. Therefore, if it is

difficult to forecast the future diesel cost, it might be prudent to not to include it as an explanatory variable. The application of 3RSSH smoothing technique was found to marginally outperform the 4253H smoothing technique. Higher forecasting accuracies were obtained for Model B. This is expected due to the use of prices for lanes which have delivery distances between 300 miles to 1500 miles which negates the impact of outliers. This result also validated the idea of splitting the data set into two categories based on distances and developing separate models. Note that for Model A which contained all delivery distances greater than 300 miles, the data set contained a number of outliers and had higher variability leading to lesser accuracies. Highest accuracies were obtained for reefer non-managed lanes. This is potentially a reflection of the fact that in the original data, reefer and non-managed lane prices had the lowest variability.

## **5.5 Conclusion**

Two types of models were developed. One where the regular time series method was applied to the raw data and the second where the data was first smoothed using robust smoothing techniques and then time series analysis methods were applied. The application of robust smoothing techniques before performing the time series analysis was found to yield better forecasting accuracies. The application of 3RSSH smoothing technique was found to marginally outperform the 4253H smoothing technique. Smoothing the datasets before applying the AR (1) process appears to be a convenient technique for avoiding irregularities or anomalies in a dataset. The variable fuel cost slightly improves forecasting results; however, univariate models developed with price-per-mile as the dependent variable and a time index (months) has shown comparable results in all cases. The models developed in this section can be easily adopted by freight and logistics companies to obtain reasonably accurate forecasts of TL prices.



## CHAPTER 6

### 6.0 EXPLORATORY SPATIAL DATA ANALYSIS (ESDA)

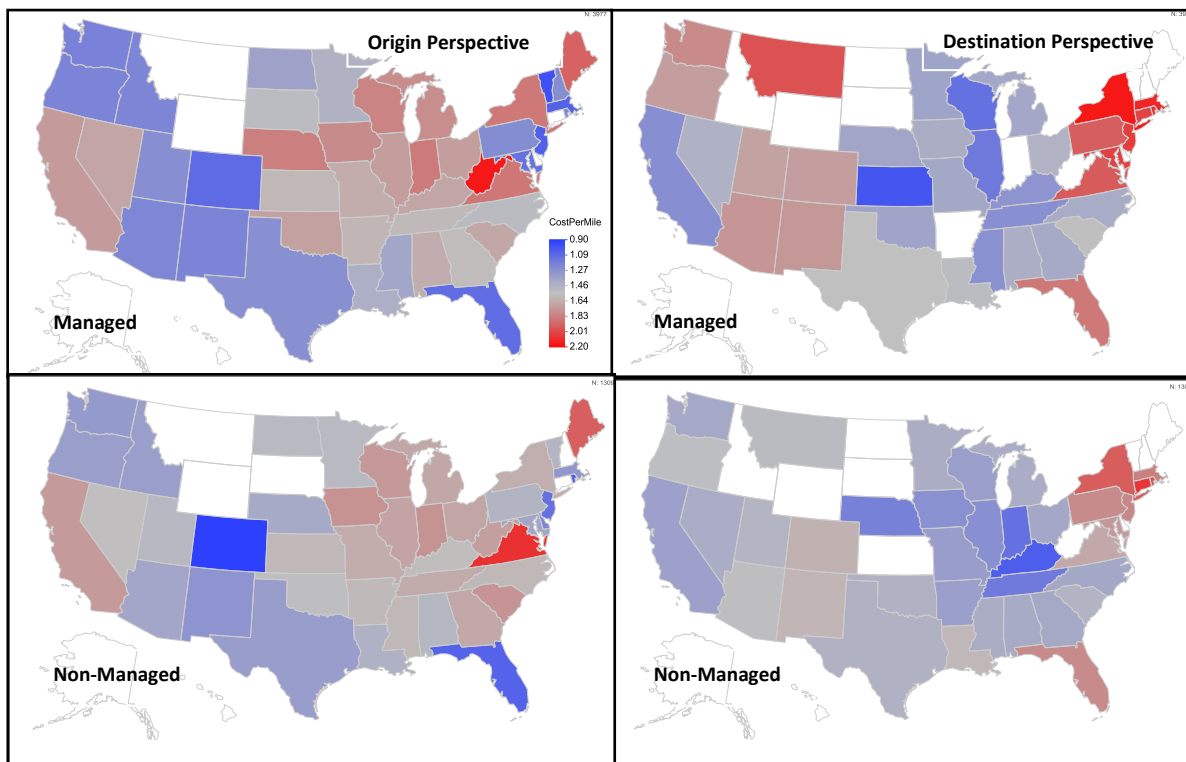
#### 6.1 Introduction

The main objective of the ESDA is to obtain insights from the dependent variable, TL price-per-mile, at the state level. The U.S. state level scale was chosen, since at lower scale visualization of maps become dense and blurry, making it difficult for human perception to capture information from the maps. U.S. states have dissimilarities such as regulations, taxes, fuel cost, number of logistic companies, number of tolls, topography, and transportation infrastructure condition. ESDA attempts to discover patterns that could be directly related to the uniqueness among locations. The illustration of large data sets in a simplified form such as a map, makes the exploratory analysis a very valuable and complex task in this research. Maps allow for a rapid interpretation of the data, but its conception relies on the ability and ideas of the creator to best represent a data set. “Maps are models of reality, by definition of models they are a simplified representation of the world” (Groff, 2006). Graphical representation of spatial data helps analysts in identifying potential clusters for modeling purposes. Significant patterns that have occurred over the last years in the truckload transportation market can be identified. This chapter performs the ESDA for the four data sets: managed, non-managed lanes and by type of commodity. Manage and non-managed data for type of commodity is described together, since alike patterns were encountered by type of commodity.

## 6.2 Exploratory Maps

### 6.2.1 Dry Loads

The data set of dry load managed lanes consists of 43,742 observations. After removing observations with delivery distance shorter than 300-mile; 39,776 observations were obtained. Dry load non-managed lanes consists of 16,332 observations. After removing observations with distance shorter than 300-mile; 13,116 observations were obtained. Data is from six years, from 2005 to 2010. Figure 11 displays the six-year averaged TL price for dry loads. The upper graphs are for managing lanes, and the lower graphs are for non-managing lanes.



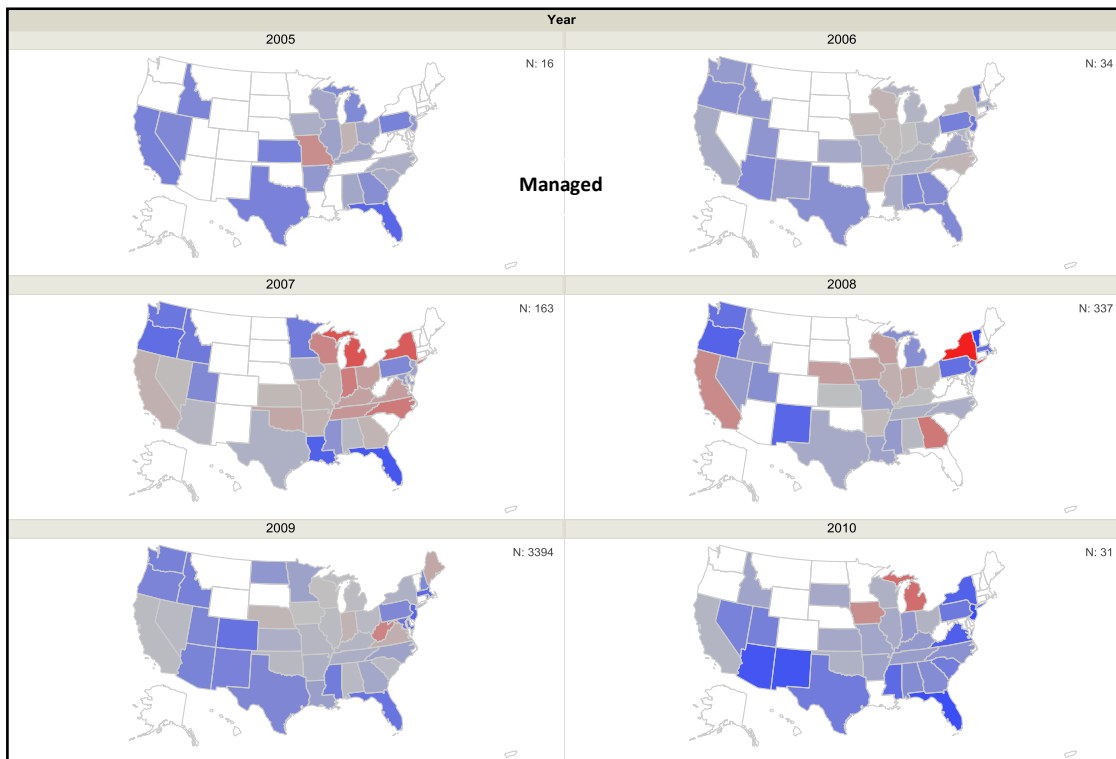
**FIGURE 11: Dry Load - Six Years Averaged**

Figure 11 shows TL prices at the origin and destination perspective. States are colored with blue, gray and red, indicating states under low, medium and high TL prices, respectively. The states

where no data was recorded are white colored. This type of notation goes throughout this chapter. From the origin perspective, data is not available on the following states; for managing lanes in: Delaware, Connecticut, Montana and Wyoming; thus, available data accounts for a total of 43 states. For non-managing lanes data is not available in: Montana, Wyoming, South Dakota, New Hampshire, and Rhode Island, available data is on 44 states. From destination perspective, data is not available on the following states; for managing lanes in: Idaho, Wyoming, North Dakota, South Dakota, West Virginia, Delaware, Maine, New Hampshire, Vermont, Arkansas, and Indiana, data available is on 37 states. For non-managing lanes in: Idaho, Wyoming, North Dakota, South Dakota, West Virginia, Delaware, Maine, New Hampshire, Vermont, and Kansas, available data is on 38 states.

Figure 11 indicates that both datasets, from the origin and destination perspective show similar patterns. From the origin perspective, high TL prices are observed in the eastern and central U.S., except for Florida and Colorado. Low TL prices are observed in western U.S. except for California and Nevada. The highest TL prices are observed in Maine, Virginia and West Virginia. The lowest TL prices are observed in Florida and Colorado. From the destination perspective, high TL prices are observed in the northeastern U.S. and Florida. Low TL prices are observed in western and central U.S. states, except for Utah, Colorado, Arizona, and New Mexico. Contrasting both maps from the origin and destination perspective, it seems that destinations with high TL prices are the origins of the deliveries with low TL prices. It is also noted that states with small area displays high TL prices from both the origin and destination perspective. In general, low TL prices are indicated in the western U.S., medium TL prices in the central U.S. and high TL prices in the eastern U.S.; the northeastern U.S has shown the highest prices, and the south central states the lowest TL prices. Figure 12 illustrates the TL price over

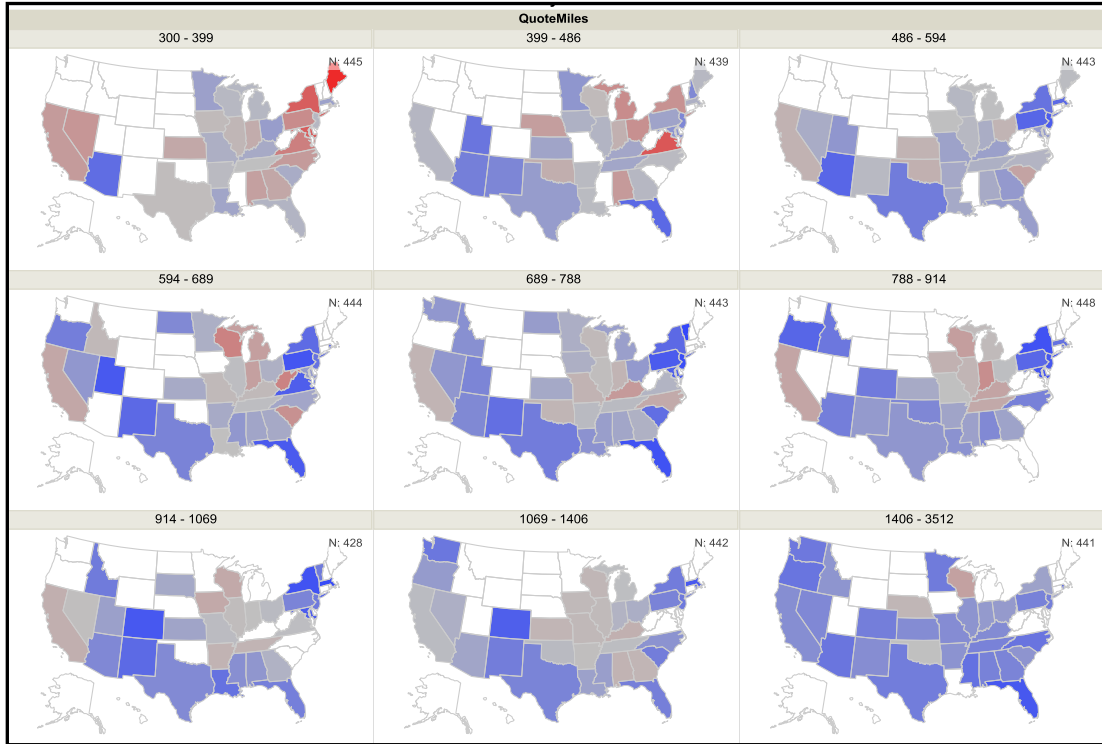
the six year period. N in the graphs indicates the number of observations in each year. Results in this analysis for managing and non-managing lanes data were encountered to have significant similarities; therefore, only the map for managing lanes data is shown in this section. The graphs for non-managing lanes are attached in Appendix A.



**FIGURE 12: Dry Load Managed & Non-Managed by Year**

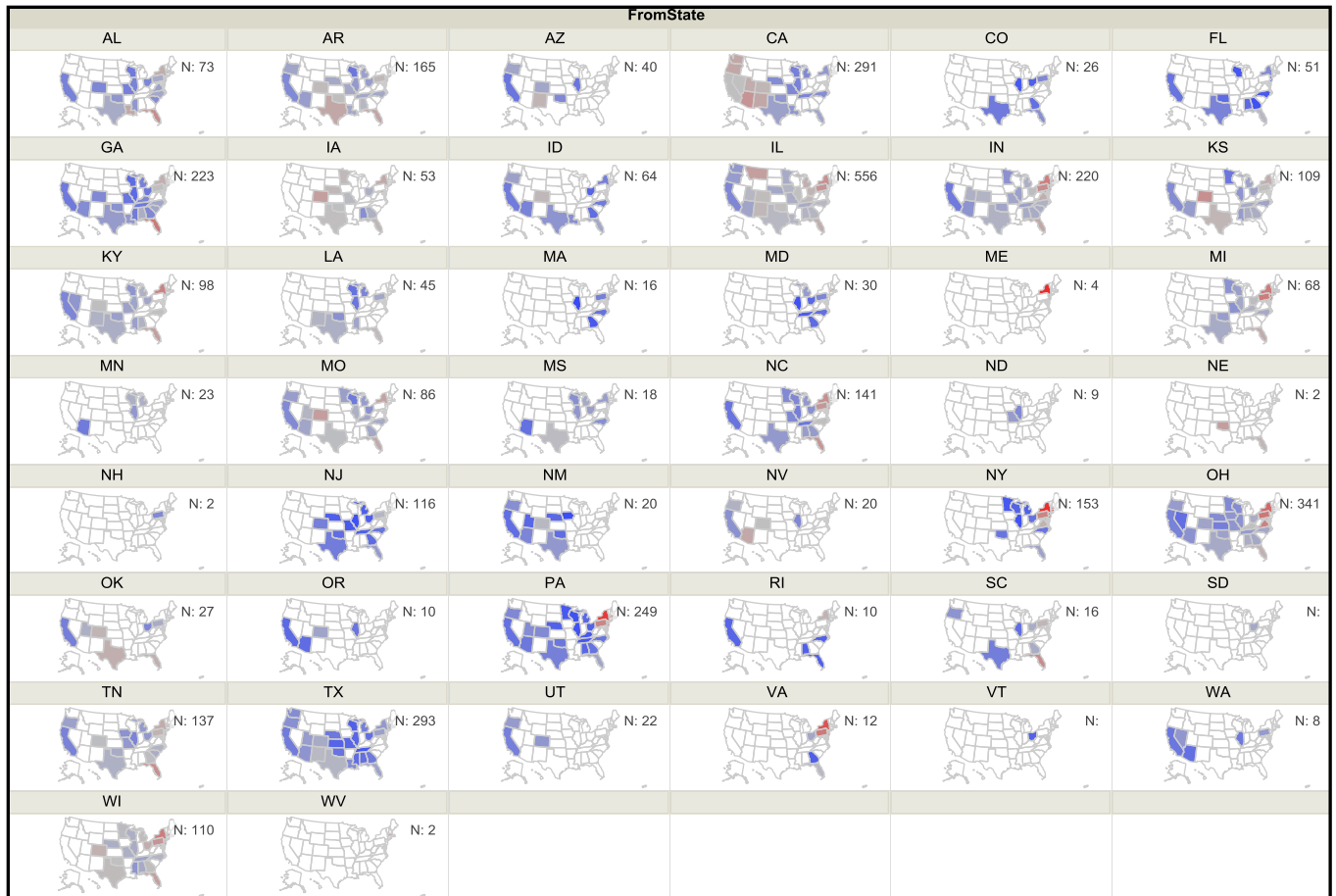
Figure 12 indicates similar patterns during the study period. In the years 2007 and 2008, TL prices were higher, mainly in the northeastern U.S. This high price may be attributed to increments of diesel cost or for more demand of commodities in those years. It is also indicated that lower U.S. states: Arizona, New Mexico, and Texas (U.S. border states of Mexico); and Louisiana, Mississippi, Alabama, and Florida have shown low and medium TL prices. It suggests that these states are less susceptible to changes of external variables affecting TL prices.

Figure 13 shows the average of TL price over the six year period divided into nine delivery distance intervals.



**FIGURE 13: Averaged TL price for Dry Load Manage**

Figure 13 shows that greater delivery distances tends to low TL prices. For short delivery distances, between 300 and 500 miles, California, Nevada, and northeastern U.S. resemble the highest prices. For delivery distances greater than 1,500 miles TL prices are steadier. U.S border state of Mexico: Arizona, New Mexico, and Texas have shown low TL prices in all delivery distances. Figure 14 illustrates the TL prices for deliveries between origin and destination states.



**FIGURE 14: Average TL Price for Dry Load Managed by O-D State**

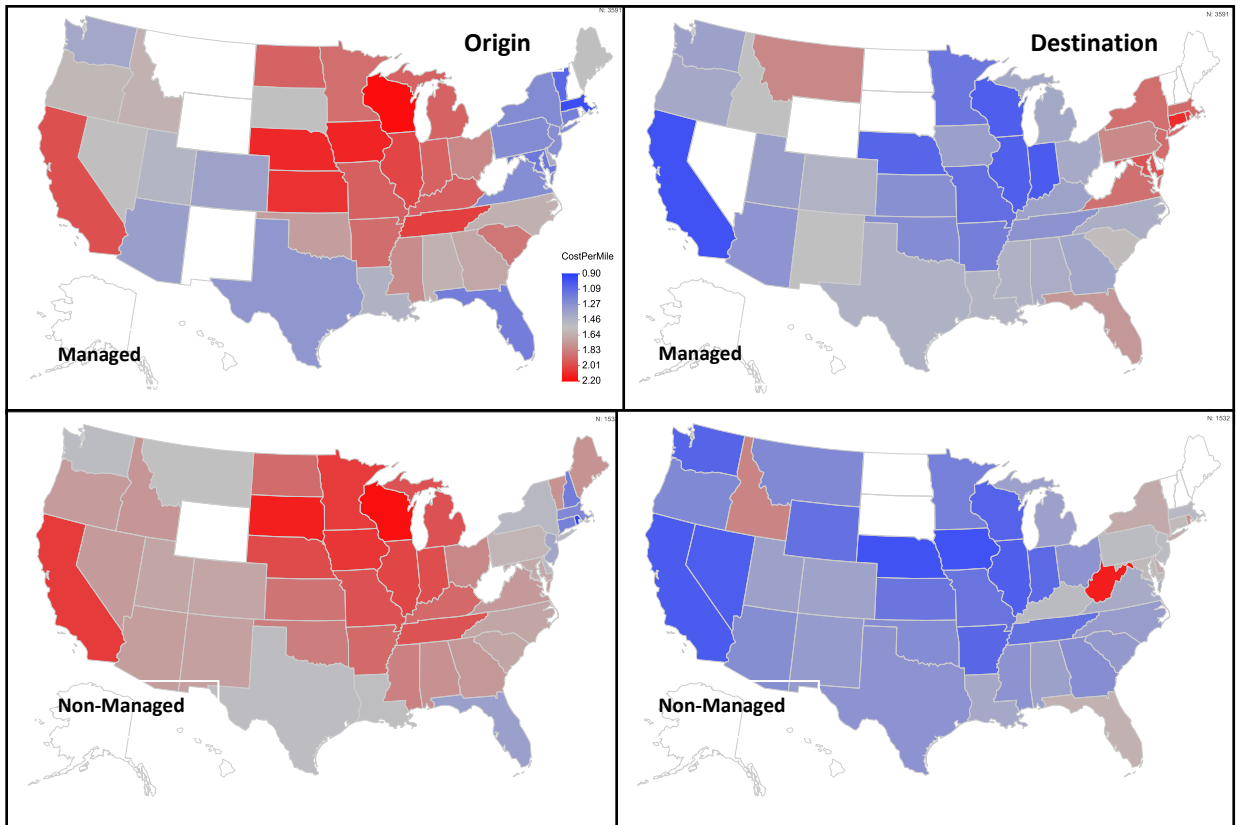
In figure 14, the first map on the upper left part, which is Alabama (AL), indicates trips from Alabama to other US. States. The perception of contiguity is provided to the reader by describing figure 14 into two categories: i) high, and low TL prices between O-D states and ii) the number of states required to cross for obtaining the destination state, which is indicated in parenthesis. The latter provides to the reader the notion of neighboring in the data description. High TL prices, as expected, are mostly related to closest states or when at least one state is crossed. The unforeseen values are for deliveries from Arizona to Florida (6), and Illinois to Montana (2) with a delivery distance greater than 1,000 miles. Low TL prices, as expected, are mostly related to deliveries that require crossing more than one state. The unforeseen values are for deliveries

from Maryland to Ohio (1), Florida to Alabama (0), Florida to Georgia (0), Georgia to Tennessee (0), Indiana to Illinois (0), New Mexico to Oklahoma (0), Alabama to Missouri (1), and from Georgia to Kentucky (1); all of them with delivery distances shorter than 600 miles.

The previous description introduces the idea of neighboring in the analysis. It provides insights regarding the relationship (weight) among states, helping for specifying an appropriate spatial weight matrix in the modeling stage. The motivation behind this analysis and for considering the number of states between the origin and destination of the deliveries is made because of the inherent effect of crossing borders as an artificial barrier of transportation trade. As described before, dissimilarities among states have an imperceptible impact to TL prices, these characteristics are complex to capture in a model; however, having knowledge of the data helps to formulate a model.

## 6.2.2 Reefer Loads

This data set consists of 37,596 observations, after removing observations with delivery distance shorter than 300-mile; 35,916 observations are obtained. This data is from six years, from 2005 to 2010. Reefer load non-managed lanes consist of 16,854 observations, after removing observations with distance shorter than 300-mile; 15,329 observations were obtained. Data is from six years, from 2005 to 2010. Figure 15 displays the six-year averaged TL price for reefer loads. The upper graphs are for managed lanes, and the lower graphs are for non-managed lanes.



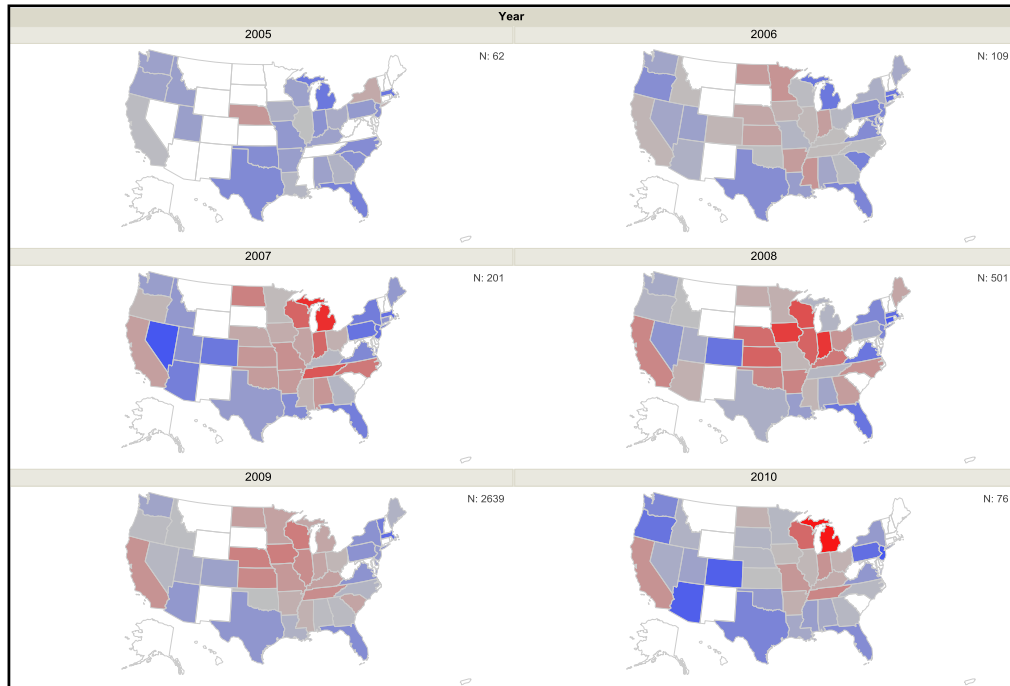
**FIGURE 15: Reefer Load Managed - Six Years Averaged**



From the origin perspective, managing lanes data is not available in the following states: Montana, Wyoming, New Mexico, West Virginia, New Hampshire, and Rhode Island. There is data available for the remaining 42 states. For non-managing lanes, data is not available in the following states: Wyoming and West Virginia; there is data available in 46 states. From the destination perspective, managing lanes data is not available in the following states: Maine, New Hampshire, Vermont, Delaware, West Virginia, North Carolina, South Carolina, Wyoming, and Nevada; there is data available in 39 states. For non-managing lanes, data is not available in the following states: Maine, New Hampshire, Vermont, North Dakota, and South Dakota; there is data available in 43 states.

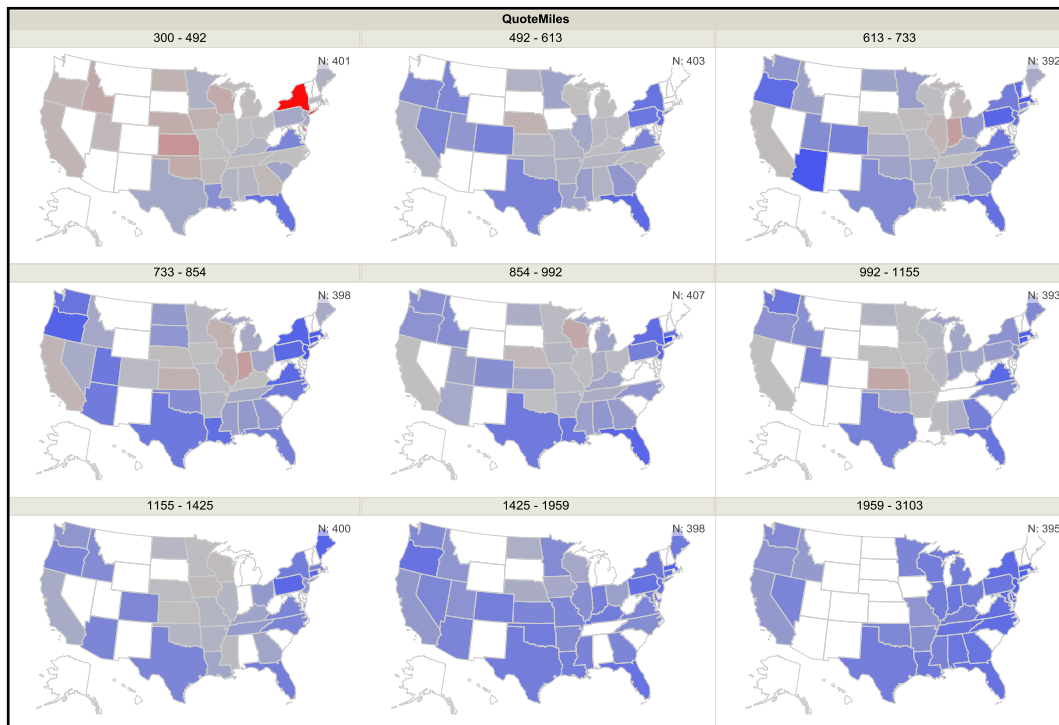
Figure 15 indicates that from the origin and destination perspective both data sets show similar patterns. From the origin perspective, high TL prices are observed in the northeast central U.S., and in California. Low TL prices are observed in northeastern U.S and in Florida. The highest TL prices are in Wisconsin, Nebraska, Iowa and California. The lowest TL prices are in Massachusetts, Rhode Island, and Connecticut. From the destination perspective, high TL prices are in the northeastern U.S. and in Florida. Low TL prices are in northeastern central U.S., and California and Nevada. The highest TL prices are in northeastern states, Florida and Montana. The lowest TL prices are shown in the northeastern central U.S. states. Contrasting both maps, from origin and destination perspective, it seems that destination with high TL prices are the origins of deliveries with low TL prices. It is also noted that states with small territory areas have low TL prices from the origin perspective. In general, from the origin perspective, low TL prices are in eastern U.S., medium and high TL prices in the western and central U.S; the northeast central U.S. has shown the highest TL prices, and the northeastern states the lowest TL prices. Figure 16 illustrates the TL price over the six year period. Results in this analysis for managing

and non-managing lanes have significant similarities; therefore, only the map for managing lanes data is shown in this section. The graphs for non-managing lanes are attached in the Appendix A.



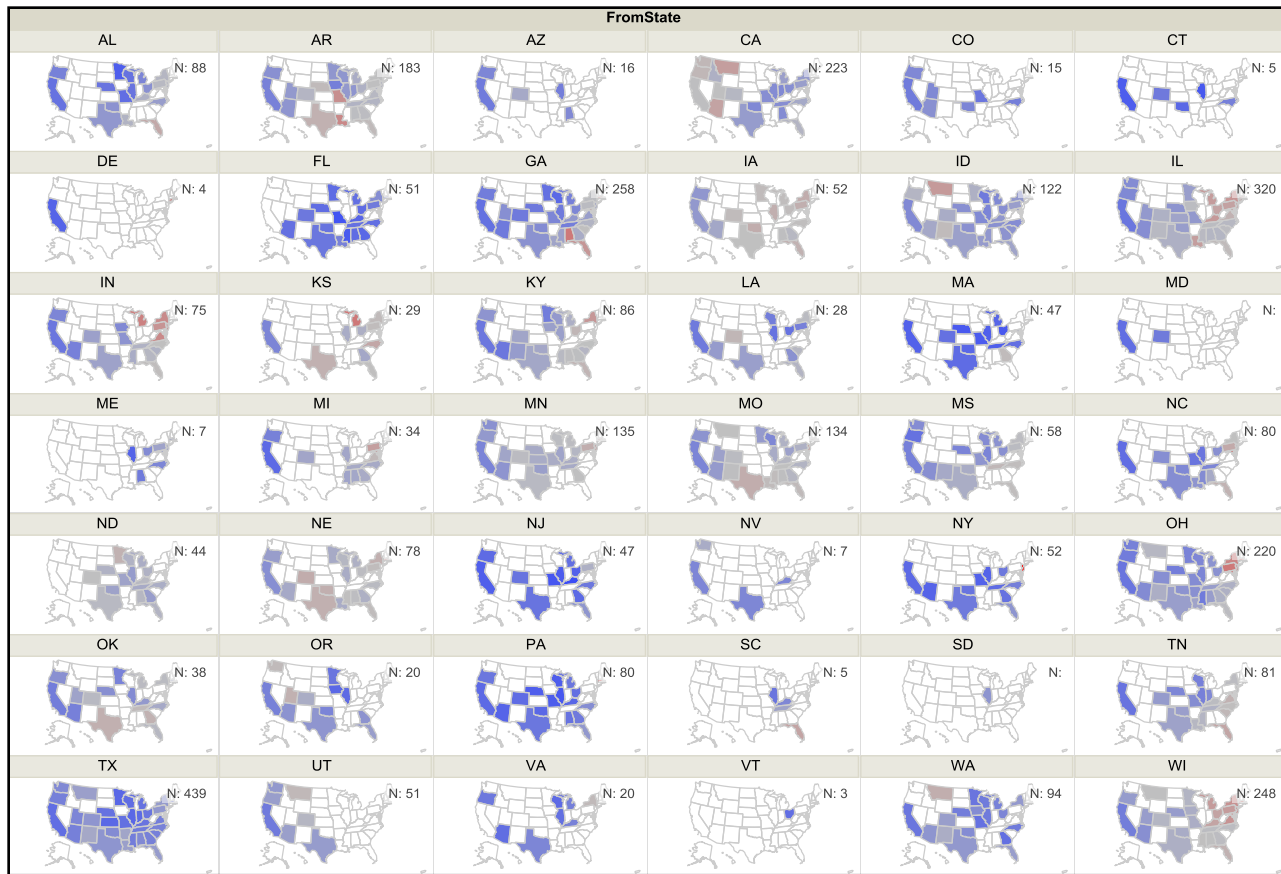
**FIGURE 16: Reefer Load Managed by Year**

Figure 16 indicates similar patterns during the study period. In the years 2007 and 2008, TL prices were higher than other years; mainly in the north central U.S. Similar to the dry loads, these high prices may attributed to higher diesel cost or more demand of goods during these years. It is also observed that lower U.S. states: Arizona, New Mexico, and Texas (U.S. border states of Mexico); Louisiana, Mississippi, Alabama, and Florida have shown low and medium TL prices. It suggests that these states are less susceptible to changes of external variables affecting TL prices. Figure 17 shows the average of TL price over the six year period divided into nine delivery distance intervals.



**FIGURE 17: Averaged TL price for Reefer Load by Distance**

Figure 17 shows that greater delivery distances tends to reduce TL prices. For short delivery distances, between 300-mile and 500-mile, New York is the only state with high prices. For delivery distances greater than 1,500-mile TL prices are low and steady. U.S. border state of Mexico: Arizona, New Mexico, and Texas have shown low TL prices in all delivery distances. Figure 18 illustrates the TL prices for deliveries between origin and destination states.

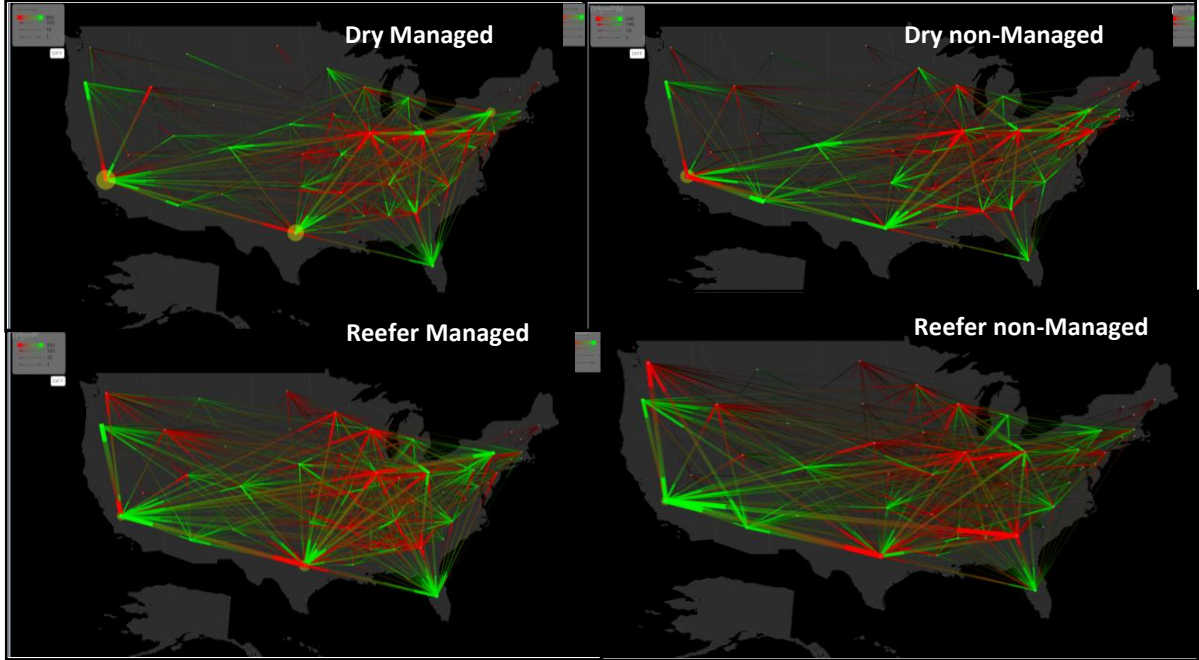


**FIGURE 18: Average TL Price for Reefer Load Managed by O-D State**

Figure 18 shows high TL prices associated with neighboring states or when one state is required to cross. The unforeseen values are for deliveries from Arizona to Louisiana (2); California to Montana (2); Illinois to New York (2), Kansas to Michigan (2); Arizona to Missouri (3); Wisconsin to New York (4); Nebraska to West Virginia (4); all of them with delivery distances greater than 1,000-mile. Low TL prices are mostly related to deliveries that require crossing more than one state. The unforeseen values are for deliveries from Pennsylvania to Ohio (0); Alabama to Missouri (1); New Jersey to Ohio (1); Alabama to Missouri (1); Arkansas to Kentucky (1); Georgia to Kentucky (1); Pennsylvania to Kentucky (1); all of them with delivery distance shorter than 600-mile.

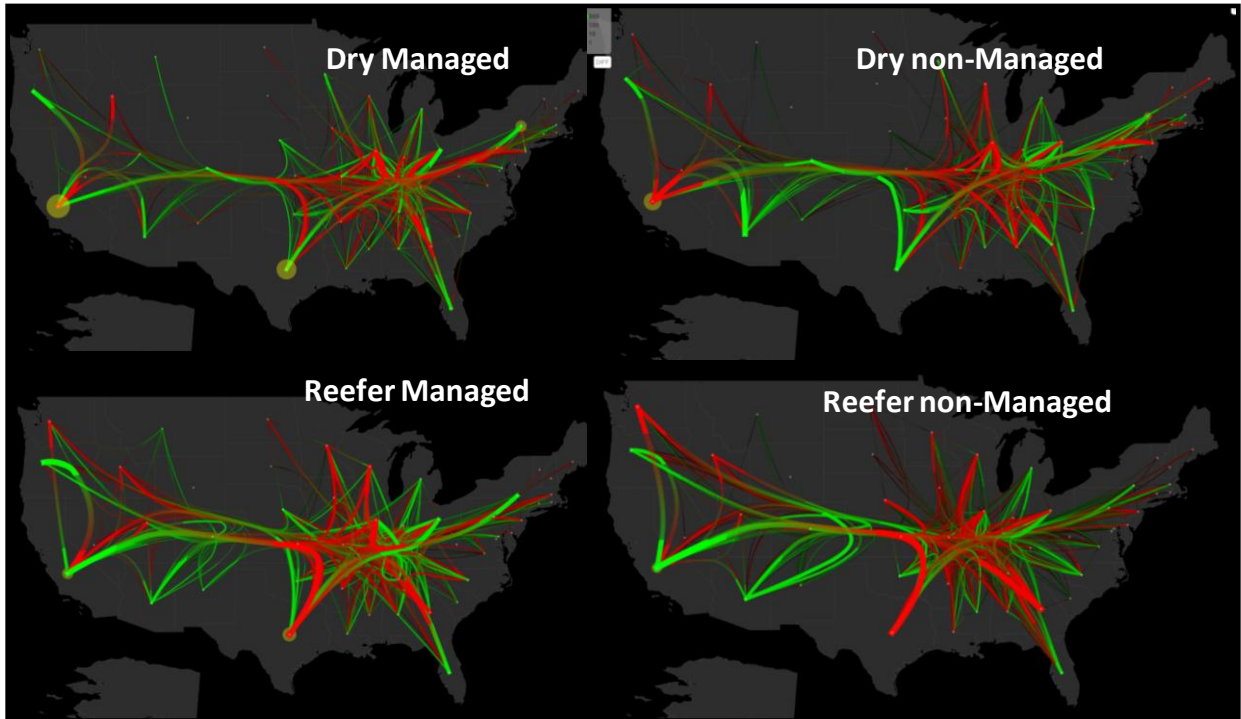
### **6.3 Interaction Maps**

Flow maps are developed for deliveries between origin and destination during each year for the four data sets. Each type of commodity has shown similar patterns for the managed and non-managed attribute; moreover, persistent patterns have been seen over the last years. The maps for each type of commodity at each year are documented in Appendix A, and the flow maps for the total number of deliveries during the six years of data are displayed in figure 19 and figure 20. Figure 19 indicates the states under light, medium and heavy volume of in-out movements; and figure 20 illustrates the bundled version of those deliveries for each type of commodity. These exploratory maps simplified and make easier for identifying major suppliers and consumer states formed by the total number of deliveries during the six year period. The reader should note that the author is reading these maps from a computer monitor, which have a better sizable image for visualization and description of data than the presented below.



**FIGURE 19: Flow Map for Total number of Deliveries**

Figure 19 shows similar flow patterns over the years for dry-managed and dry-non managed, as well as, for the reefer managed and non- managed data. For dry load, it is observed that California, Florida and Texas are large consumer states, mainly from the eastern and central U.S. For reefer load is observed that California, and Florida are large consumer, mainly from the central and northeastern U.S. For both types of commodities, California is the major supplier of Oregon, and Illinois is shown as the major supplier of California, Texas, New York, and Florida. U.S. states serving as local suppliers for their neighboring states are Illinois, Kansas, and Arkansas.



**FIGURE 20: Bundle Map for Total Deliveries in the U.S.**

Figure 20 shows California, Florida, Texas and New York as the large dense areas in both types of commodities, which indicates major trade zones in the U.S. Strong accumulation of hosts are observed in the bundle map, on the eastern and western U.S, which suggests three remarkable regions based on truckload movements: the eastern, central and western U.S. Busy trade directions are also observed from the eastern to the western U.S. states, and between states from the lower to upper eastern U.S. states. From figure 19 and 20 significant patterns for each type of commodity is described below.

#### Dry loads

- Major suppliers are: Illinois, New Jersey, Pennsylvania, Kansas, Kentucky, New Hampshire, Vermont, Maine, North Dakota, and South Dakota, California, Wisconsin, Ohio, Maine, Rhode Island, Idaho, and Washington.

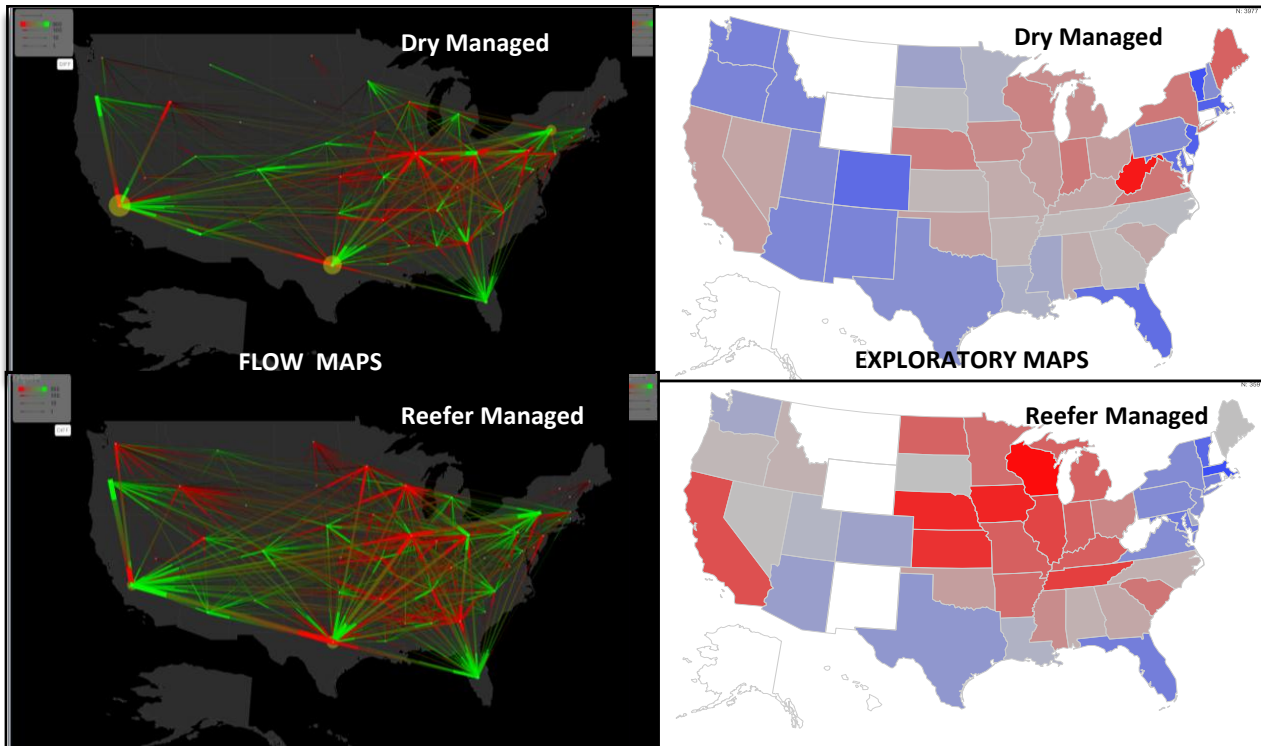
- Major consumer states are: Florida, New York, Minnesota, Oregon, Colorado, Arizona, Nebraska, Michigan, Massachusetts, Montana, Texas, and North Caroline.
- States with large volume of incoming and out coming flows are: California, Texas, Georgia, Ohio, Missouri, Illinois, Minnesota, Wisconsin, and Alabama.
- Large volume of deliveries are indicated from California to Oregon, Texas, and Arizona; from Texas to California, Florida, Colorado; from Illinois to California, Texas, New York, Florida and Minnesota; from Pennsylvania to New York, California, and Texas.

#### Reefer Loads

- Major suppliers are: Texas, Georgia, Illinois, Arkansas, Wisconsin, Minnesota, North Dakota, and Idaho.
- Major consumer states are: California, Florida, Colorado, Nebraska, Montana, and North Caroline.
- States with large volume of incoming and out coming flows are: New York, Texas and Ohio.
- Large volume of deliveries are indicated from California to Oregon; from Texas to California, Florida, Arizona, and New York; from Illinois to California, Texas, New York, and Florida; from Minnesota to Texas, California and Illinois; from Idaho to surrounding states, from Wisconsin to surrounding states.

Figure 21 illustrates the exploratory TL price maps and the flow maps to observe patterns regarding TL price and the state's attribute of being a producer or a consumer state.





**FIGURE 21: Flow and Exploratory Map**

Figure 21 shows the flow maps for dry and reefer managed lane data, the exploratory map on the right hand is based on the origin perspective. Figure 21 indicates that for dry and reefer loads, consumer states are most likely to have low TL prices. For instance, Florida, Colorado, Utah, Arizona, New Mexico has shown persistent low TL prices through the analysis. High TL prices are mostly associated with producer states, for instance, it is observed in the north central U.S., where maps are mostly colored red. States that have shown dissimilar patterns are California and Nevada, consumer states with high TL prices.

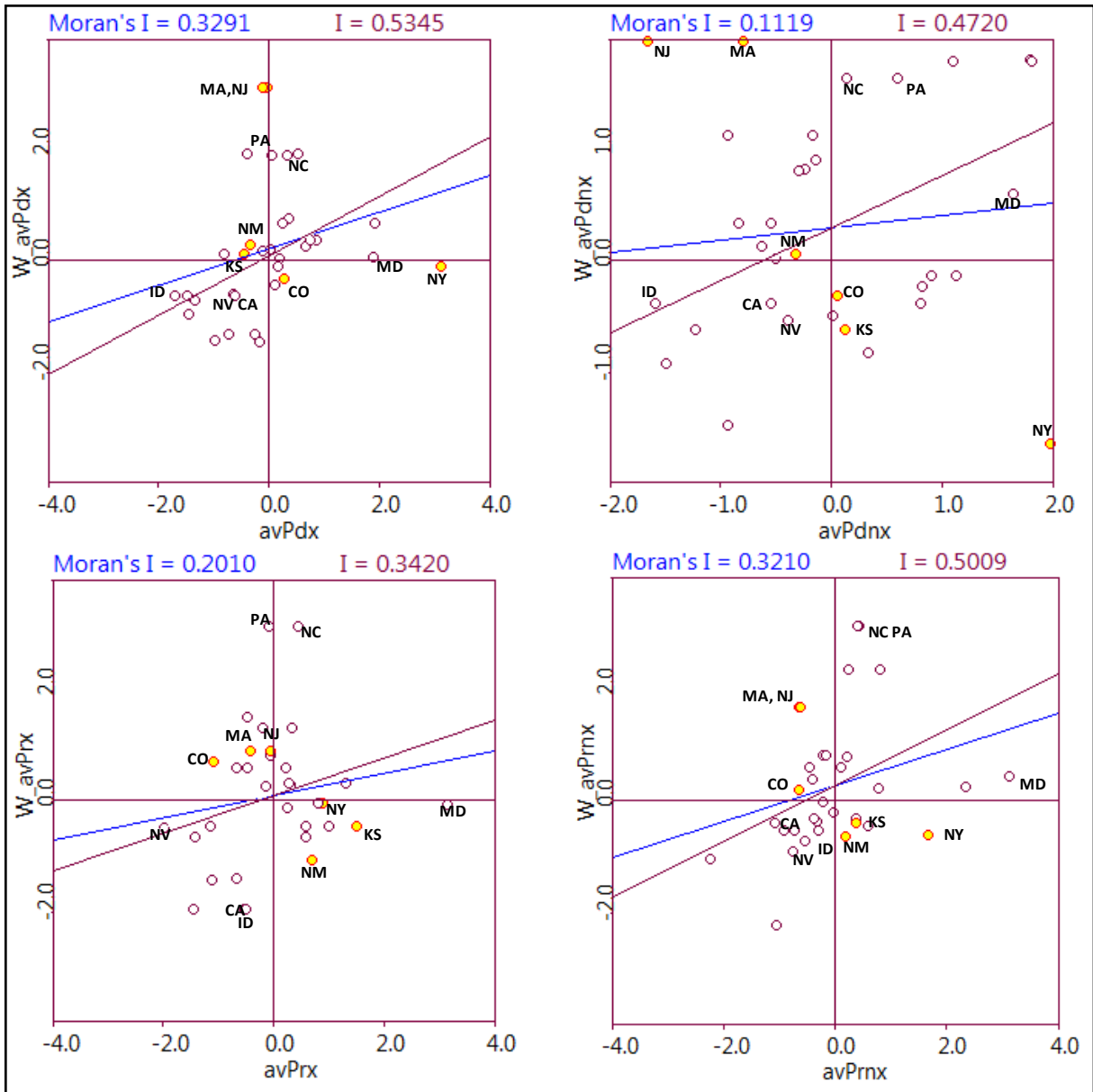
## 6.4 Spatial Autocorrelation

Spatial autocorrelation analysis was evaluated throughout the Moran's I test for each data set. From previous analysis, a notion regarding to an appropriate SWM for the datasets was assessed. The main characteristic in all data sets is the relationship between delivery distance and TL price, larger distance are associated with lower TL prices. Therefore, neighboring proximity, as well as, distance between neighbors has shown as a feasible alternative for developing an appropriate SWM.

The first nearest-neighbor SWM was considered in this analysis. It considers the matrix  $w_{ij}=1$ , if "j" is one of the first nearest neighbors, and 0 otherwise. This type of SWM is usually used when spatial interaction is associated with neighbor proximity (Lee, 2009). Other SWM specifications such as queen, rook, and contiguity were also tested, showing comparable results; therefore, only results for first-nearest neighbor is reported in this section.

For each year spatial autocorrelation was evaluated (from 2005 to 2010) to investigate the behavior of each state over the years, by observing its position on the Moran's I scatter plot at each year. Then, each state was assigned to one of the four categories: HH, LL, LH, and HL (quadrant position in the scatter plot), on the basis of its likelihood of being in a specific location in the scatter plot. The same analysis was performed, but now using the average of TL price for the six years of the study period. Results from both analyses have shown comparable results in terms of the location of each state in the scatter plot. Therefore, the analysis using the average of TL price was conducted for the univariate plot, and including the inverse distance variable for the bivariate plot.

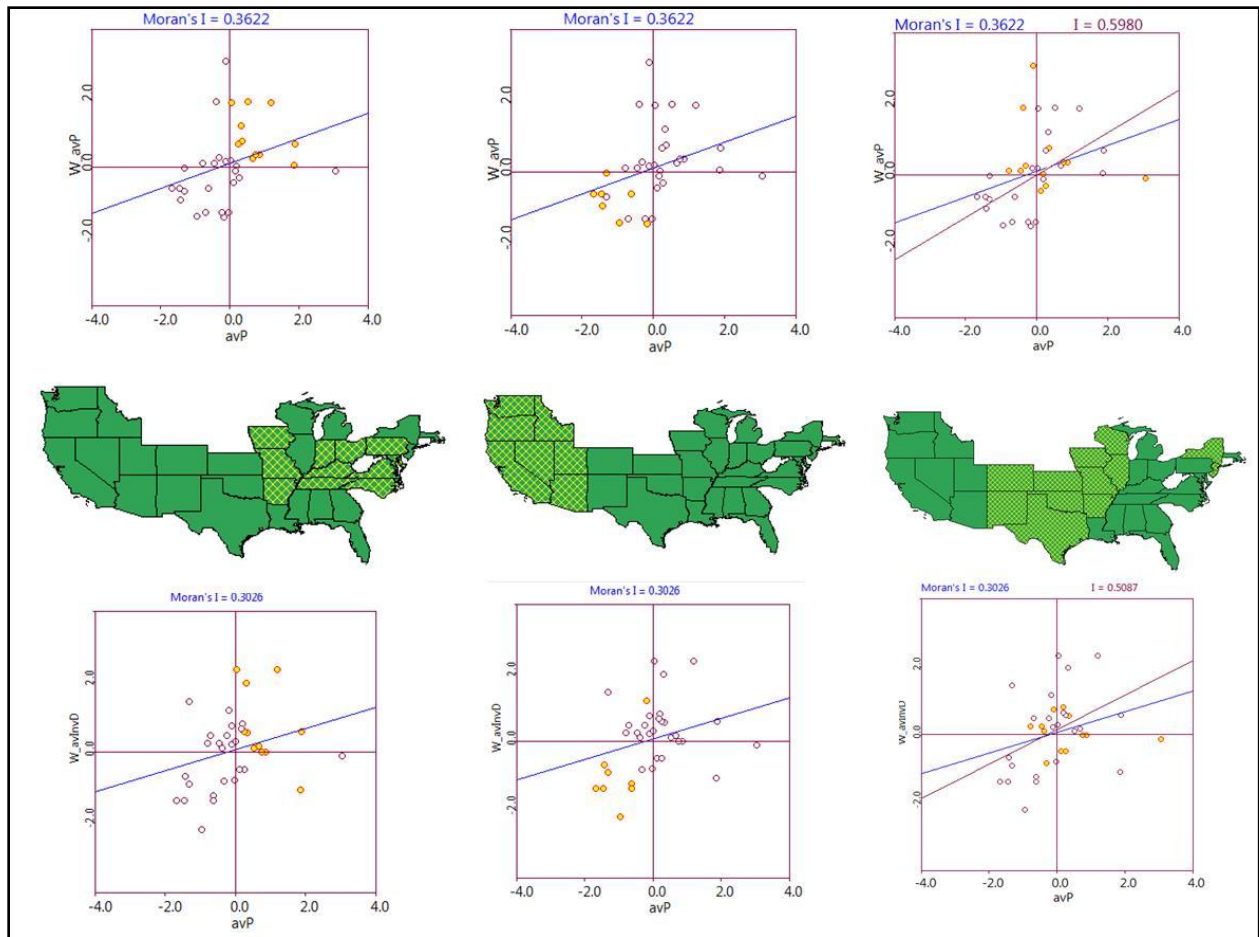
Figure 22 illustrates the univariate Moran's I scatter plot for each data set. It is important to note that these scatter plots were developed using the available states in the four datasets, thus, a missing state in one data set was deleted in the others.



**FIGURE 22: Moran's I Scatter Plot for each Dataset**

Figure 22 shows positive spatial autocorrelation for the four data sets. For illustration purpose, a few states were highlighted in the scatter plot to perceive resemblances among the four data sets by observing their location on each scatter plot. For instance is observed that Maryland, Pennsylvania, and North Carolina are mostly located in quadrant I (High-High); and Idaho, Nevada and California in quadrant III (Low-Low) for all data sets. These two groups of states have shown positive spatial autocorrelation. Similarly, New Jersey, Massachusetts and New Mexico are in the quadrant II (Low-High), and New York, Colorado and Kansas in the quadrant IV (High-Low) in all the data sets. These last groups of states showed negative spatial autocorrelation.

Figure 23 illustrates the univariate and bivariate Moran's I scatter plot for dry managed lanes data. This analysis includes all the states with data available. Comparable results, and persistent patterns were found for the other data sets; thus, only dry managed lane data are reported. The univariate plot was developed by using the average TL price-per-mile (upper plots), and bivariate plots was developed with the addition of the variable delivery distance in its inverse form (lower plots), a regression line on each plot is also shown. The first nearest-neighbors SWM was used in the analysis. It was created using a Geoda tool, setting  $k=1$ . Figure 23 also displays brushing values as yellow points in the Moran's I scatter plots, which are reproduced on the map as hatched states.



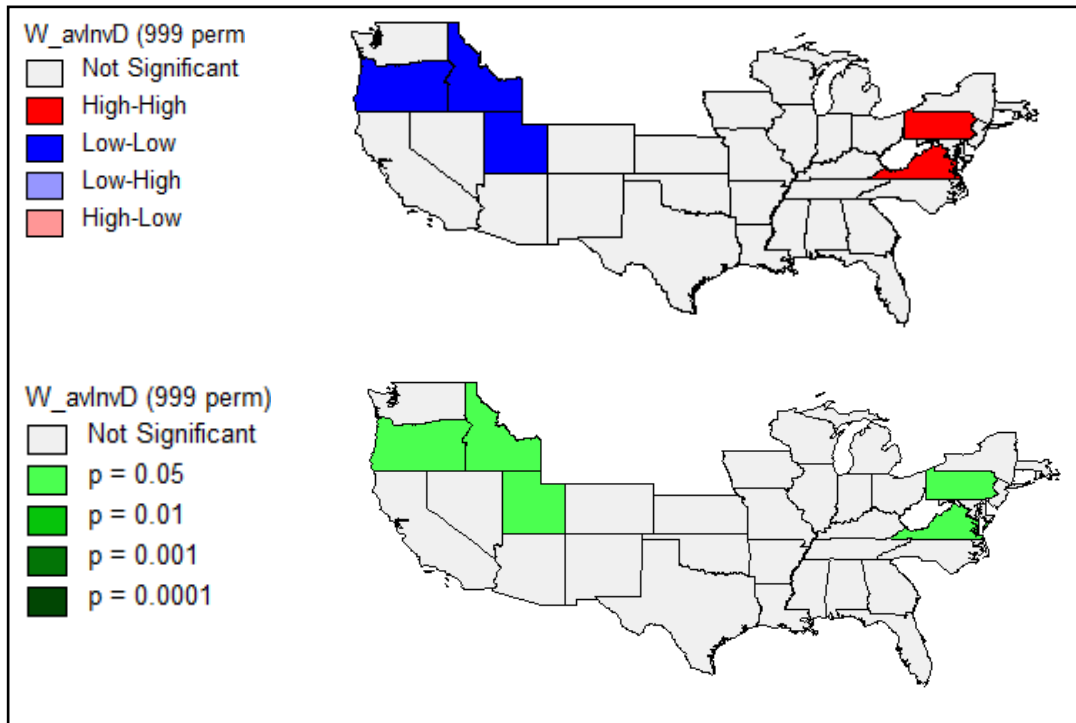
**FIGURE 23: Moran's I test: Univariate and Bivariate Scatter Plot**

Figure 23 displays that most of the states fall into quadrants I (High-High) and III (Low-Low), which indicate a positive spatial autocorrelation. States with positive correlation, quadrant I and III are associated with the eastern and the western U.S, respectively. States showing negative spatial autocorrelation, quadrant II and IV, are associated with the central U.S. As previously mentioned, comparable results were found in the four data sets. A description regarding the location of each state in the Moran's I scatter plot is described below.

- Quadrant I: Iowa, Pennsylvania, Indiana, Ohio, Maryland, Kentucky, Virginia, Missouri, North Carolina, Tennessee, Arkansas, and Michigan.

- Quadrant II: New Jersey, Illinois, Kansas, Texas, New Mexico, and Georgia
- Quadrant III: Washington, Idaho, Oregon, Nevada, Utah, California, Arizona, Oklahoma, Alabama, Mississippi, Louisiana, and Rhode Island.
- Quadrant IV: New York, Colorado, Oklahoma, Florida, and Massachusetts

It is observed that most of the states in quadrant I and III belong to the eastern and western U.S. (positive spatial autocorrelation), and quadrant II, and IV to the central U.S. (negative spatial autocorrelation). Also, the location of Kansas and Massachusetts differs from the one obtained in figure 23; however, in both, they have shown negative autocorrelation. In addition, GeoDa allows users to brush for outliers and automatically recalculated the Moran's I value. This process was conducted by brushing states on the quadrant II, and quadrant IV, and it is shown on the scatter plots in the right hand of figure 23. It is observed that by excluding these states from the analysis, the Moran's I test goes from 0.362 to 0.598 in the univariate analysis, and in the bivariate analysis from 0.302 to 0.509, indicating strong positive autocorrelation. The previous analysis evidences three main regions: the east, west, and central U.S. It is confirmed by observing the Lisa multivariate plots illustrated in figure 24.



**FIGURE 24: LISA Cluster and Significance Map**

Figure 24 illustrates the LISA cluster and significance map. The randomization was applied for 999 random values to obtain the significance at 5% of confidence level. The randomization is conducted to generate results with some degree of stability. It indicates that Oregon, Idaho, and Utah have low TL prices and are surrounding for states with low TL price; Pennsylvania and Virginia, which have high TL prices are surrounding to states with high TL price. These states are hatched in the graphs since they are significant at 5% of confident level. This analysis also evidences and confirm that surrounding states at those locations (red and blue) have similar high and low TL prices, which are not by random chance, they have some degree of spatial autocorrelation.

## 6.5 Conclusion

Significant patterns are described in this analysis for dry and reefer commodities in the U.S territory at the state scale level. For dry commodities, high TL prices are observed in the eastern U.S; the northeastern states exhibit the highest prices. Medium and low TL prices are observed in the central and western U.S., respectively. The lowest TL prices are in the southwest U.S. The exploratory analysis for reefer commodities indicates that high TL prices are in the central U.S.; the north-central U.S. exhibits the highest TL prices, medium and low TL prices are in the western and eastern U.S. Particular zones: eastern, central and western are identified from the exploratory maps associated with TL prices.

The TL prices for dry and reefer commodities indicate that high prices are related to deliveries between neighboring states, or when one state separates the origin and the destination. Low TL prices are associated with deliveries when the origin and destination are separated for more than one state. Lower U.S. states such as the U.S. states border of Mexico have shown low and steady TL prices over the study period. The territory area of the state reveals a positive relationship with TL price; small size states are most likely to high TL prices from the origin perspective. The last statement also could indicate that low density states, usually low consumers, are likely to have high TL prices. Moreover, destinations with high TL price are likely to be the origins of deliveries with low TL prices. Major trade zones in the U.S. are identified in California, Florida, Texas and New York; these zones are under a large volume of deliveries as shown in the flow maps.

Consumer and producer states were identified for both commodities, from the origin perspective, consumer states are mostly associated with low TL prices, and producer states are associated with high TL prices. Strong accumulation of movements in the eastern and western



U.S is identified, similar to the exploratory maps, flow maps reveal three zones in the U.S. with high, medium and low TL prices in the central, western and eastern U.S. The Moran's I test and Lisa plots of TL prices has revealed strong evidence of spatial autocorrelation, the majority of states falls into quadrants I and III, which indicates a positive association. It also evidences similar remarkable regions in the U.S. as shown in the other exploratory maps; the eastern, central and western U.S.

The analysis also has shown insight regarding state under the disfavorable conditions for truckload deliveries. For dry commodities, Arizona and Illinois, as the origin of the deliveries, exhibit the major constraints or barriers for truckload deliveries, they present higher TL prices even in suitable situations such as large delivery distances. Maryland, Florida, Georgia, Indiana, New Mexico, and Alabama, as the origin of the deliveries, exhibit favorable conditions for truckload deliveries, they have shown low TL prices even in unfavorable situations such as short delivery distance. For reefer commodities, Arizona, California, Illinois, Kansas, Wisconsin, Nebraska, as the origin of the deliveries, exhibit the major constraints for truckload deliveries; and Pennsylvania, Alabama, New Jersey, Arkansas, Georgia, and Pennsylvania, as the origin of the deliveries, exhibit favorable conditions for truckload deliveries.

The analysis also indicates specific characteristics for some states. For instance, Florida and Oregon appear as an exclusive consumer state for both commodities with Texas and California as their major supplier states, respectively. California, as expected, resembles the largest consumer in the U.S. for both types of commodities with Texas and Illinois as its major supplier. California, Florida, Texas and New York shows the highest incoming and outgoing volumes of deliveries, indicating major areas of domestic trades in the U.S. Texas and Illinois have exhibited as the major producer states for both commodities.

The three regions found in this analysis can also be subdivided into sub-regions such as the north, south and mid sub-region; thus, the similarities encountered in their TL prices can be aggregated. These regions are shown below.

### **East Region**

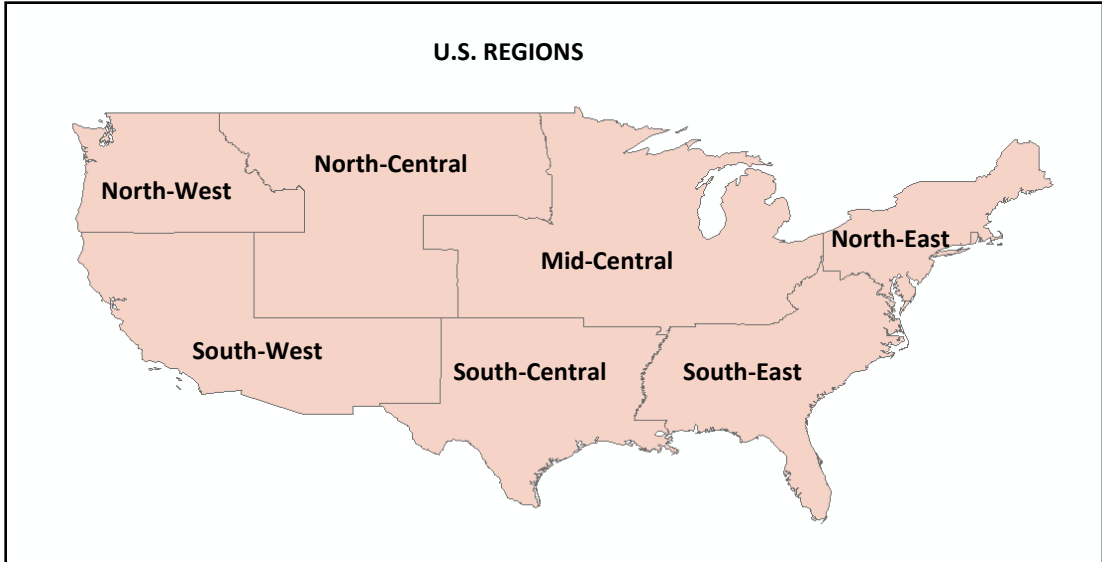
- North-East (NE): Connecticut, Delaware, Massachusetts, Maine, New Jersey, New York, Pennsylvania, Rhode Island, Vermont;
- South-East (SE): Alabama, Florida, Georgia, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia;

### **Central Region**

- North-Central (NC): Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming;
- South-Central (SC): Arizona, Louisiana, Oklahoma, Texas;
- Mid-Central (MC): Iowa, Illinois, Indiana, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, Ohio, Wisconsin;

### **West Region**

- North-West (NW): Arkansas, Idaho, Oregon, Washington;
- South-West (W): Arizona, California, New Mexico, Nevada.



**FIGURE 25: U.S. Regional Map Associated with TL Prices**

## CHAPTER 7.0

### 7.0 SPATIAL ANALYSIS

#### 7.1 Introduction

Spatial econometric analysis was conducted for the database. The literature review section described the motivations and important relationship between the transportation activity and geography, and their combined effects to the society. This relationship is confirmed in the EDA and ESDA section, where TL prices and geographical space are investigated. Therefore, this association should be considered in the analysis. The main objective of this chapter is developing a model at the state scale, which incorporates geographic features, allowing to draw inferences throughout explanatory variables, and to forecast TL prices considering different scenarios. Conventional approaches are only focused on capturing the linear relationship between the dependent and explanatory variables, and fail to adequately address the space effects. For instance, inadequate transportation infrastructure conditions and deficient connectivity between locations could hamper freight movements. It generates congested roads, delay shipping, add cost and unreliability to the deliveries. The proposed model differs from traditional models, because it does not provide a single output based on the inputs. It allows for creating different scenarios across space. The analyst can test different inputs under various modeling parameters, and visualize into a map. A valuable feature of the proposed model is its capability of measuring the direct and indirect impacts across units. The proposed model can calculate the impact on the dependent variable in a specific location, when changing an independent variable at one specific location.

In the process of formulating a model, various spatial specifications documented in the spatial econometrics literature were considered. The TL price database consists of cross-sectional

data collected at different locations and periods of time. The dependent variable is the yearly average of TL prices for each state, considering the origin state of the deliveries. Hence, the present analysis is at the state scale level with TL prices of the origin of the deliveries from 2005 to 2010 as the dependent variables. The spatial panel models include a SWM which was selected by ESDA insights, and by testing various SWM specifications. The selected SWM has the ability of producing expected signs of coefficient estimates, and acceptable forecast values in the four data sets.

In addition, the Lagrange multiplier and Hausman test are used, the first test for evaluating spatial autocorrelation either in the dependent variable or in the residuals, and the latter test for helping in selecting between fixed and random effects in the spatial model. This section begins by testing the data sets for spatial autocorrelation in the dependent variable or in the residuals, which helps in selecting an appropriate SAR or SEM specification. Then, models are tested between the fixed or random effects. The most significant effect is incorporated in the model based on the Hausman test. After obtaining the appropriate model specification and effect, different explanatory variables were tested to obtain the final model. The final model is used for estimating the direct and indirect effects across locations, as well as, for forecasting TL price values for the year 2010. The MAPEs values are calculated and reported.

## **7.2 Database**

The data used in this analysis are a mix of private data (truckload rates) and publicly available datasets (socioeconomic variables). The database for this analysis consists in TL prices for deliveries between states. TL prices are for deliveries between the 48 contiguous U.S. states; however, not all states have available data in the four data sets. The socioeconomic variables

tested in this analysis are gross domestic product (GDP), population, unemployment rates, hourly truck wages, and personal income. The previous variables were obtained in the Bureau of Economic Analysis ([www.bea.gov](http://www.bea.gov)) and the truck driver wages in the Bureau of Labor Statistics ([www.bls.gov](http://www.bls.gov)).

### **7.3 Spatial Weight Matrix**

The selected SWM is the first-nearest neighbor, considering the 48 contiguous U.S. states. The SWM are standardized to have row sums of unity, which allows for producing linear combinations of observations from neighbors in the model.

### **7.4 Spatial Autocorrelation**

Spatial autocorrelation is conceptually more difficult to understand than time series autocorrelation, mainly because the ordering assumption is not held (Dubin, 1998). Usually, time series assumes that earlier observations affect later observations, this assumption allow researchers for testing and developing time series models. This ordering assumption is not valid for cross-sectional data and autocorrelation among them, because whether unit 1 effects unit 2 is possible that the reverse happens and also affects neighboring observations. The Moran's I test conducted in a previous chapter confirmed the spatial autocorrelation in the data sets, but the association of this autocorrelation with either the dependent variable or the residuals is not established. The Lagrange multiplier test helps in determining the most appropriate structure of the model. The residuals are tested for evaluating their significance under SAR or SEM models. The test for the most significant effects, between fixed and random effects is reported in the next section.

### 7.4.1 The Lagrange Multiplier Test

The Lagrange Multiplier (LM) test is the major diagnostic tests for spatial dependence either in the dependent variable or in the residuals. The test is estimated by maximum likelihood, assuming normality for the perturbations. The LM lag test and the LM error test evaluate spatial autocorrelation in the dependent variable and in the residuals, respectively. These tests do not consider the presence of the alternative form of spatial dependence e.g. autocorrelation in both, the dependent variable and the residuals. Therefore, robust LM tests are also conducted to observe any form of dependence (Anselin et al, 1996). Florax, Folmer and Rey (2003) indicates that when the null hypothesis of no spatial autocorrelation is rejected, additional tests which evaluate spatial correlation in the residuals in presence of spatial lag or spatial correlation in the dependent variable in the presence of spatial error dependence should be considered. LM test results for each data set is displayed in table 5.

**TABLE 5: Lagrange Multiplier Test for Spatial Panel Data**

	<b>Dry Managed</b>	<b>Dry non-Managed</b>	<b>Reefer Managed</b>	<b>Reefer non-Managed</b>
<b>LM lag test for omitted spatial lag</b>				
LM value	20.783	32.089	7.597	30.049
Marginal Probability	0.000	0.000	0.006	0.000
Chi (1) .01 value	6.640	6.640	6.640	6.640
<b>LM error test for spatial correlation in residuals</b>				
LM value	17.085	26.121	6.206	22.992
Marginal Probability	0.000	0.000	0.013	0.000
Chi (1) .01 value	6.640	6.640	6.640	6.64
<b>Robust LM lag test for omitted spatial lag</b>				
LM value	6.711	8.947	8.873	11.597

Marginal Probability	0.009	0.003	0.003	0.000
Chi (1) .01 value	6.640	6.640	6.640	6.640
<b>Robust LM error test for spatial correlation in residuals</b>				
LM value	3.014	2.978	7.482	4.540
Marginal Probability	0.083	0.084	0.006	0.033
Chi (1) .01 value	6.640	6.640	6.640	6.640

Results indicate that the LM-Lag, Robust LM-Lag, and LM-Error tests are significant at 5% of confident level for the four data sets. However, the Robust LM-Error test is only significant for the reefer managed data sets. It is also noted that the LM values for the Lag tests are higher than the Error tests in all the data sets. These tests suggest lags in the dependent variable or SAR models for all data sets

### 7.5 Spatial Econometric Model Specification

Spatial econometric models were conducted for each data set. The models were developed under the spatial econometric panel model theory. The main objective is capturing the spatial effects in the model. The spatial econometric specification is designed with the incorporation of the spatial autocorrelation parameters and the selected SWM. The spatial lag model is the basis formulation with the variables in their log form. Various model's versions are developed by considering a set of explanatory variables, and considering dependency across space and time. The final model selected is one that provides the expected sign and reasonable coefficient estimates, which provides the most accurate forecast values.



### 7.5.1 Spatial Autoregressive Model (SAR)

The spatial econometric model selected is a spatial-lag model, which have spatial dependency in the dependent variable as shown in equation 26. This type of model was considered due to the four data sets falls under this specification according to table 5. However, additional analysis under the SEM specification were also tested and reported in this section.

$$y_t = \alpha\iota_N + distance_t\beta + unempl_t\theta + \rho W y_t + e; \quad (26)$$

$$e \sim N(0, \sigma^2 In)$$

The variables in the models are in their log forms. The dependent variable,  $y_t$  represents the TL price at period of time  $t= 2005...2010$ . The significant explanatory variables in the model are delivery distance and unemployment. The parameter  $\alpha\iota_N$  represents either the random or fixed intercept in the spatial panel model,  $\rho$  is the spatial autoregressive parameter,  $\beta$  and  $\theta$  are the coefficients associate with the explanatory variables; delivery distance and unemployment, respectively; and  $W$  is the SWM.

### 7.5.2 Fixed v.s. Random Effects (Hausman Test)

The intercept term  $\alpha\iota_N$  in the model capture the effects of omitted variables. These effects could be considered as a constant effect across units, called fixed effects or act randomly called random effect. The random effect model, consider a stochastic variable that can be attributed to a known distribution, and that variable is uncorrelated with the explanatory variables. The fixed effects consider a constant but random variable, which is allowed to be correlated with the explanatory variables. Differences of opinions and consideration between the uses of these two effects in a model are described in the literature review. The Hausman test helps analyst to evaluate which effect, fixed or random, is suitable for the model by calculating the likelihood of

each effect in the model. This computation is performed by comparing coefficient estimates from two models, each one with one particular effect. The likelihood is provided by the t-statistic and p-value. A MATLAB routine developed by Elhorst (2010) was used in the analysis, results for each data set is shown in table 6.

**TABLE 6: Hausman Test between Fixed and Random Effects**

	<b>T-Statistic</b>	<b>Degree of Freedom</b>	<b>Probability</b>
<b>Dry managed</b>	70.6284	2	0.000
<b>Dry non-managed</b>	449.1351	2	0.000
<b>Reefer managed</b>	41.286	2	0.000
<b>Reefer non-managed</b>	6.330	2	0.042

Table 6 indicates that at 5% of confident level dry data sets and reefer managed data are associated with fixed effects, a constant parameter for each unit. Reefer non-managed lanes fall into the random effects at 5% of confident level, at 10% of confident level, all data sets falls into the fixed effect hypothesis. It is important to highlight that this test uses a one-sided chi-squared curve, which means that probability lower than 0.025 involve rejection of the random effects (null hypothesis) in favor to the fixed effects. As previously described, this test helps analysts to discern between the fixed and random effect in the model; however, the inclusion of one or the other effect is by the discretion and judgment of the researcher as indicated in the literature part of this research.

### 7.5.3 Parameter Estimation

The estimation of parameters in the models considers the maximum likelihood method as described by Elhorst (2010). A preliminary analysis was conducted using all the explanatory variables in the study: delivery distances, unemployment, fuel cost, population, truck driver wage, and gross domestic product. This analysis considers the spatial panel model of the SAR, SDM, SEM and SDEM specification, which are documented in the Appendices B and C.

Results indicate that SAR random and fixed effects produce similar coefficient estimates for the significant variables: delivery distance and unemployment in the four data sets. Delivery distances have a greater impact in the managed data than in the non-managed data, while unemployment behaves in the opposite direction, a greater impact to the non-managed data in dry loads. In the reefer loads delivery distance and unemployment have greater impact in the non-managed lanes data. Unemployment is not a significant variable for the reefer managed data, being fuel cost, a statistically significant variable for this dataset. In all the models the coefficients present the expected signs. The coefficients from the SAR models are larger than the ones in the SEM model, but both keep the same expected signs. The intercept is only significant in the SEM model with random effects.

As a result from the preliminary analysis, delivery distances seem statistically significant in all datasets, while unemployment is for all, except for reefer managed lanes data, where fuel cost shows of being significant. The spatial Durbin models (SDM, SDEM) do not display any improvement in the models based on comparing the standard deviation of the models. SAR and SEM show better results. Comparing SAR and SEM models, between the random and fixed effects, shows similar results; however the SAR model provides better coefficient estimates based on the t-statistics and p-values. The SAR, SEM, SDM, SDEM, with fixed and random

effects were developed using these explanatory variables for the four data sets. Table 7 illustrates the SAR model with spatial fixed effects and table 8 with random effects.

**TABLE 7: Spatial Panel Autoregressive Models (SAR\_F) with Spatial Fixed Effects**

Parameter	Dry Managed	Dry non-Managed	Reefer Managed	Reefer non-Managed
R-squared	0.2767	0.2429	0.2668	0.2049
corr-squared	0.1616	0.3215	0.1360	0.2378
$\sigma^2$	0.0288	0.0331	0.0175	0.0185
$\alpha$	--	--	--	--
$\beta$ ( <i>dist</i> )	-0.255 (0.000)	-0.19794 (0.000)	-0.19198 (0.000)	-0.22943 (0.000)
$\theta$ ( <i>unempl</i> )	-0.1820 (0.000)	-0.3662 (0.000)	-0.02357 (0.4264)	-0.209338 (0.000)
$\rho$ ( <i>W.yt</i> )	-0.2361 (0.000)	-0.2361 (0.000)	-0.2361 (0.000)	-0.2361 (0.000)

p-value in parenthesis

Effects	Dry Managed	Dry non-Managed	Reefer Managed	Reefer non-Managed
Direct	-0.2633 (-5.4763)* -0.1892 (-4.9202)*	-0.2043 (-3.6351)* -0.3829 (-8.3689)*	-0.1978 (-5.7592)* -0.0241 (-0.7548)*	-0.2375 (-5.1774)* -0.2175 (-7.0955)*
Indirect	0.0576 (3.5617)* 0.0416 (3.2248)*	0.0454 (2.8848)* 0.0857 (3.8004)*	0.0427 (3.5159)* -0.0052 (0.7258)	0.0514 (3.5833)* 0.0473 (3.7947)*
Total	-0.2057 (-5.3656)* -0.1476 (-4.9914)*	-0.1588 (-3.6019)* -0.2972 (-9.0331)*	-0.1551 (-5.7973)* -0.0189 (-0.7545)	-0.1861 (-5.0811)* -0.1702 (-7.4155)*

t-statistics in parenthesis; \*1% significance; \*\* 5% significance; \*\*\* 10% significance

Table 7 displays coefficient estimates significantly different from zero and with the expected signs. Results indicate that greater delivery distances tend to reduce the TL price in terms of dollars-per-mile, similarly higher unemployment rates tend to reduce prices. It suggests that in low demand market conditions, carriers are likely to provide discount in their prices. It is also observed that unemployment is not significant for the reefer managed lanes data. An additional analysis has shown fuel cost as significant as seen in Appendix B. The elasticity of the delivery

distance variable for the four data sets drives from 0.198 (reefer managed) to 0.264 (dry managed). The elasticity of the unemployment rate for the four data sets drives from 0.189 (dry managed) to 0.383 (dry non-managed); reefer non-managed lanes have a elasticity of 0.2175. It is important to remark that the direct effects are slightly different from the coefficient estimates ( $\beta$ , and  $\theta$ ) due to the feedback effects that affects neighbor locations and return to the own location. Both explanatory variables; delivery distance and unemployment show positive indirect effects. The spatial autoregressive parameter ( $\rho$ ) is statistically significant, negative and relatively large, -0.236, which indicates that changes in TL prices in neighboring location impacts in the opposite direction to the owning location. It suggests that greater delivery distance tends to reduce TL price in the own state, but it attempts to raise TL prices in neighboring states.

The logical interpretation of the last statement is explained in the scenario of having a particular trucking company, which usually provides services without the restriction of providing services to short or large delivery distances. Therefore, under low demanding conditions, the lack of services is expected, reducing the production of vehicle-mile-traveled-per-year for the carrier. Therefore, this carrier is anticipated to reduce its TL price by prioritizing large delivery distance, since it makes possible a reduction of prices accordingly as shown in the model. In this scenario, neighboring carriers are most likely to serve the short and medium delivery distances, which have higher TL prices. A similar interpretation can be provided in the high unemployment rate scenario, because high unemployment rates will reduce the consumption of products, reducing the production of transportation services. Thus, carriers are more likely to implement strategies such as prioritize services to large deliveries distances. The results of the spatial lagged model with random effect are reported in table 8.

**TABLE 8: Spatial Panel Autoregressive Models (SAR\_R) with Spatial Random Effects**

<b>Parameter</b>	<b>Dry Managed</b>	<b>Dry non-Managed</b>	<b>Reefer Managed</b>	<b>Reefer non-Managed</b>
R-squared	0.1299	0.0820	0.1199	0.0302
corr-squared	0.1747	0.2487	0.1503	0.2154
$\sigma^2$	0.0288	0.0334	0.0175	0.0188
$\alpha$ (effect)	2.4692 (0.000)	2.4929 (0.000)	2.0229 (0.000)	2.4078 (0.000)
$\beta$ ( <i>dist</i> )	-0.2475 (0.000)	-0.2117 (0.000)	-0.2013 (0.000)	-0.22456 (0.000)
$\theta$ ( <i>unempl</i> )	-0.1574 (0.000)	-0.2901 (0.000)	-0.0294 (0.257)	-0.151453 (0.000)
$\rho$ ( <i>W.yt</i> )	-0.2361 (0.000)	-0.2361 (0.000)	-0.2361 (0.000)	-0.2361 (0.000)
theta	0.7307 (0.000)	0.9973 (0.000)	0.7974 (0.000)	0.9972 (0.000)

p-value in parenthesis

<b>Effects</b>	<b>Dry Managed</b>	<b>Dry non-Managed</b>	<b>Reefer Managed</b>	<b>Reefer non-Managed</b>
Direct	-0.2559 (-6.6008)* -0.1645 (-4.6231)*	-0.2173 (-4.9696)* -0.3006 (-7.8574)*	-0.2062 (-7.288)* -0.0297 (-1.099)	-0.2309 (-6.625)* -0.1556 (-6.046)*
Indirect	0.0556 (3.8637)* 0.0359 (3.0882)*	0.0478 (3.5049)* 0.0665 (3.8597)*	0.0444 (3.930)* -0.0063 (1.058)	0.0496 (3.794)* 0.0335 (3.611)*
Total	-0.2003 (-6.4752)* -0.1285 (-4.7743)*	-0.1695 (-4.8865)* -0.2340 (-8.2775)*	-0.1619 (-7.261)* -0.0234 (-1.095)	-0.1813 (-6.652)* -0.1221 (-6.182)*

t-statistics in parenthesis; \*1% significance; \*\* 5% significance; \*\*\* 10% significance

Table 8 displays coefficient estimates significantly different from zero and with the expected signs. The coefficients are similar to the ones obtained for the fixed effects analysis (table 7). Comparing results between fixed and random effects, the coefficient associate to the delivery distance variable indicates a 4% variation, and 20% for the unemployment rate variable. Theta is the parameter of the log-likelihood function. It is important to point out that the random effects are drawn from a normal distribution.

**TABLE 9: Spatial Error Models (SEM\_F) with Spatial Fixed Effects**

<b>Parameter</b>	<b>Dry Managed</b>	<b>Dry non-Managed</b>	<b>Reefer Managed</b>	<b>Reefer non-Managed</b>
R-squared	0.4077	0.4177	0.4135	0.3781
corr-squared	0.2023	0.3419	0.1601	0.2595
$\sigma^2$	0.0213	0.0204	0.0128	0.0119
$\alpha$	--	--	--	--
$\beta$ ( <i>dist</i> )	-0.14913 (0.000)	-0.155625 (0.000)	-0.15016 (0.000)	-0.174866 (0.000)
$\theta$ ( <i>unempl</i> )	-0.127585 (0.000)	-0.2590 (0.000)	-0.011568 (0.7022)	-0.143518 (0.000)
$\rho$ ( <i>W.yt</i> )	0.2139 (0.000)	0.298995 (0.000)	0.196985 (0.000)	0.2989 (0.000)

p-value in parenthesis

Table 9 displays coefficient estimates significantly different from zero and with the expected signs. The coefficient estimates are smaller than the ones obtained in the SAR\_F models. A similar interpretation of the coefficients can be provided. Similar to the previous analyses, unemployment is not significant for reefer managed data. The elasticity of the delivery distance variable for the four data drives from 0.149 (dry managed) to 0.175 (reefer non-managed). The elasticity of the unemployment rate for the four data sets drives from 0.127 (dry managed) to 0.259 (dry non-managed); elasticity of reefer non-managed lanes is 0.143. Both variables show positive indirect effects. Similar interpretation as the previous analysis can be considered. The spatial autoregressive parameter ( $\rho$ ) is statistically significant, positive and relatively large, between 0.20 and 0.29, which indicates that changes in TL prices in neighboring location impacts in the same direction to the owning location. These models lack of measuring the indirect impacts to neighboring locations. This model indicates that reduced prices in one state influences prices attempt to reduce prices in other states, which is anticipated in a competitive market. This model lacks of providing the spillover impacts of price changes to

neighboring locations, a more general interpretation of the results. Table 10 reports the results of the error model with random effect.

**TABLE 10: Spatial Error Models (SEM\_R) with Random Effects**

<b>Parameter</b>	<b>Dry Managed</b>	<b>Dry non-Managed</b>	<b>Reefer Managed</b>	<b>Reefer non-Managed</b>
R-squared	0.3538	0.4455	0.3576	0.3896
corr-squared	0.2053	0.2798	0.1774	0.2416
$\sigma^2$	0.0214	0.0202	0.0128	0.0118
$\alpha$	1.8425 (0.000)	1.9937 (0.000)	1.6294 (0.000)	1.8165 (0.000)
$\beta$ ( <i>dist</i> )	-0.179030 (0.000)	-0.181988 (0.000)	-0.15809 (0.000)	-0.16455 (0.000)
$\theta$ ( <i>unempl</i> )	-0.115677 (0.000)	-0.183508 (0.000)	-0.037644 (0.130147)	-0.11556 (0.000)
$\rho$ ( <i>W.yt</i> )	0.21598 (0.000)	0.301221 (0.000)	0.1871 (0.000)	0.307185 (0.000)
<i>teta</i>	0.0553 (0.2003)	0.056011 (0.1837)	0.27622 (0.000)	0.116628 (0.0204)

p-value in parenthesis

Table 10 displays coefficient estimates significantly different from zero and with the expected signs. The coefficients associated with distance are smaller than the one obtained with the SEM\_F, but higher estimates associated with unemployment. The variation in the coefficient estimates between the fixed and random effects is about 20% in delivery distance, and 17% for unemployment. Theta is only significant for the reefer data sets.



#### 7.5.4 Forecasting Results

The main objective of this analysis is forecasted short term TL prices, in this case one year ahead. Similar to the non-spatial model, the year 2010 was held for validating purpose. It is important to mention that for calibrating the spatial models in this process not consider data for the 2010 year. MAPE values are shown in table 11.

**TABLE 11: Accuracy of State Level TL Price Forecasting**

<b>DATA</b>	<b>SAR_F MAPE</b>	<b>SAR_R MAPE</b>	<b>SEM_F MAPE</b>	<b>SEM_R MAPE</b>
<b>Dry Managed</b>	18.79%	16.50%	18.48%	18.62%
<b>Dry non-Managed</b>	18.15%	17.07%	19.62%	18.89%
<b>Reefer Managed</b>	10.69%	11.06%	13.19%	10.48%
<b>Reefer non-Managed</b>	7.66%	8.59%	10.11%	9.56%

Table 11 indicates good forecast results with MAPE values less than 20% for all models. The highest accuracy is for the reefer non-managed data with MAPE values less than 11%, highly accurate forecasts. The SAR model with fixed and random effects seems that provides the lowest MAPE values.

#### 7.6 Summary and Conclusions

A spatial analysis was conducted testing various explanatory variables: population, state gross domestic product, fuel cost, driver wages, and unemployment. The t-statistics indicates delivery distance and unemployment as the significant variables in all data sets, except for reefer managed lane, where fuel cost is significant. The coefficient estimates are not highly affected when considering either fixed or random effects in all data sets. Results indicate that greater

delivery distances as well as higher unemployment rates tend to reduce the TL prices in terms of dollars-per-mile. In low demand market conditions, the analysis suggests that carriers are more likely to reduce their prices, which sounds reasonable. Results from the SAR with fixed effects provides the most significant estimates and also good forecast outcomes. SAR model indicates that for a 1.00 percent change of delivery distance the TL price is affected in about -0.25 percent, and for a 1.00 percent change in unemployment the TL price is affected in -0.30 percent. It suggests that unemployment has greater impact than the delivery distance variable to TL price. Positive signs for the direct effects were obtain for delivery distance and unemployment. The indirect effects indicates that for 1.00 changes in delivery distance or unemployment tends to reduce TL prices in neighboring regions in about +0.05 percent and +0.06 percent respectively. Results suggests that greater delivery distance reduce prices in the own location, but tends to raise prices in neighboring locations due to the spillover effects. The forecasting models provide good forecasts with MAPE values less than 20% for all models. The reefer non-managed lane data have the highest accuracy with MAPE values less than 11%.

## CHAPTER 8

### 8.0 CONCLUSION AND FUTURE WORK

#### 8.1 Findings

This research uses original methods for approaching and analyzing truckload prices in the U.S. Exploratory analyses from the non-spatial and spatial perspective have been conducted and evaluated, which allows for a superior visual exploration and interpretation of data, enhancing modeling decisions during the analysis. The efficient use and operation of these tools take part in the contribution of this research. The spatial analysis incorporates the unemployment rate variable in the model as a proxy of the consumer demand for each state. This analysis indicates that TL prices are closely associated with the demand of services that links economic carriers' market conditions. Consequently, demand takes an important role when setting prices. Results indicate and validate that splitting the data into two categories, according to delivery distances, is convenient for forecasting purpose. Separate models, considering distance greater than 300 miles, reduce the impact of outliers in the datasets and the higher variability leading to lesser accuracies. TL rates between \$10.00 and \$30.00 per mile are observed when trucking companies work under 300 miles delivery distances as shown in the EDA chapter. Models developed by the time series techniques can be easily adopted by freight and logistics companies to obtain reasonably accurate TL price forecasts.

The exploratory spatial data shows high, medium and low TL prices in the eastern, central and western U.S. for dry loads; the highest prices are in the northeastern, and the lowest in the southwest U.S. For reefer loads: high, medium and low TL prices are in the central, western and eastern U.S., respectively; the highest prices are in the north-central U.S. The analysis reveals the relationship between territory area and TL prices. Small states have high

prices, and large states usually are associated with low TL prices. Major trade zones are California, Florida, Texas and New York; strong accumulation of incoming and outgoing volumes of deliveries is in these states as shown in the flow maps. Florida and Oregon are large consumers of goods for dry and reefer loads. Major suppliers are Texas, Illinois and California. California is the largest consumer in the U.S. as was expected. Flow maps also display strong accumulation of movements in the eastern and western U.S. Busy trade directions is perceived from the eastern to the western U.S. states, and between states from the lower to upper eastern U.S. states as shown in the bundle maps. The three types of visualization maps; exploratory maps, flow maps, and spatial autocorrelation maps evidence three remarkable U.S. zones, the eastern, western and central U.S. These zones can also be divided into seven regions as shown in figure 24.

Another important finding is based on how I described the data, which considers location that provides to the reader with the sense of proximity, in this case, the number of states required to cross between the origin state and the destination state. This analysis finds the routes with more barriers, which have higher prices than other routes with similar distances. By this method, origin and destination states have shown disfavor conditions for truckload deliveries. For instance, when Arizona and Illinois are the origin of the deliveries for dry loads, these states exhibit the major constraints or barriers for truckload deliveries with the highest TL prices.

Delivery distance is the most significant variable associated with TL prices in all models. The unemployment rate variable is statistically significant in all datasets, except for reefer-managed lane. The coefficients estimates of the spatial models are not substantially affected when considering between fixed and random effects in all models. The Hausman test favors the fixed effects for all data sets, except for the reefer non-managed lanes. However, fixed effect is

preferred since it provides similar or better results than the random effect in all models. The spatial model displays that larger delivery distances and higher unemployment rates implicate lower TL prices. Therefore, during low demand market conditions such as high unemployment rate, carriers are likely to reduce prices, and they are most likely to serve larger delivery distance to reduce cost and to be more competitive. The inferences analysis indicates that +1.00% increase in delivery distance tends to decrease TL price by about -0.25%; this drop affects neighboring locations by increasing TL price by about +0.05%. For instance, if we consider the average delivery distance of 1,500 miles and the average TL price of \$1.55-per-mile, a 150 miles increase in delivery distance tends to decrease the prices-per-mile by about 0.039 cents, and this discount tends to raise prices in neighboring locations by about 0.008 cents. Moreover, the analysis suggests a +1.00% change in the unemployment rate tends to decrease TL price by about -0.30%, and this will increase the price of neighboring locations by about +0.06%. In other words, serving larger delivery distance tends to decrease TL price in the immediate location, but it attempts to raise TL prices in neighboring locations.

Forecasting methods conducted in this research has shown acceptable accuracy results. A striking finding in the non-spatial and spatial analysis is that diesel cost was not a significant variable for TL price setting during the last years. The economic market condition seems to be the major issue when prices are setting. Robust methods resistant to outliers have shown outstanding results in the non-spatial models for forecasting TL prices in the U.S. MAPE values were less than 10% in all cases. The spatial lagged model with fixed effects provides better forecasting performance than other models. MAPE values were less than 20% in all cases. The lowest variability is in the reefer non-managed data, which shows the highest accuracies in both models.

## **8.2 Discussion**

The negative effects of high unemployment rates on consumer demand and their impacts on the economy of a region are well understood. In the U.S., unemployment rates have been increasing since 2006 (BLS, 2012), altering the U.S. economy, and the consumer behavior associated with demand for services and products. These changes have impacted TL patterns in comparison with previous years. Usually, people reduce spending on non-essential services and on long-lasting manufactured products such as vehicles, furniture, electronics, and also vacations. Saving money and reducing debt is the priority of people. High unemployment rates drop the demand of goods, mainly non-essential dry products, because reefer products are considered most essential for people. The drop in consumption reduces production of services for trucking companies, usually calculated in terms of vehicle-mile-traveled-per-year, impacting TL rates as has been shown in this research. In a low demand scenario, carriers are willing to offer price discounts seeking strategies for being more competitive and less economically impacted. This price reduction allows carriers to attract more services, to keep their vehicles out of storage, and to enable to cover their fixed costs e.g. vehicle insurances, license, loans, etc. The particular question of how TL carriers will respond to high or low consumer demand is important for logistic companies, policy makers and transportation planners. This question is addressed in this research, and also exposes this price reduction and its impact on neighboring locations.

## **8.3 Contribution of Dissertation**

This study is the first to introduce, displays and document truckload price patterns in the U.S. It would be considered as a starting point for understanding truckload patterns related to prices and movements across the U.S. An original truckload price database in the U.S. was

investigated in this research. This research empirically proves the relationship of TL prices with consumer spending throughout the unemployment rate variable as a proxy of consumer demand. This work provides various perspectives that expand the understanding of patterns, incorporating important and original information to the literature, and also potential methodologies and insights for forecasting truckload prices in the U.S. The use of private and publicly available data makes this research particularly important, since it allows for integrating data from different sources. The use of exploratory data techniques and tools, and how to handle and display large data sets in the transportation field is also a contribution of this research.

Original methods, which have not been tested in the freight transportation domain provides valuable results. These methods allow for obtaining data insights, for adjusting the spatial and time dependence, for drawing significant inferences from the data, and for providing accurate forecast TL prices. The information documented in this research will allow practitioners in the freight transportation field for visual interpretation of the truckload market related to prices and movements in the U.S. It provides the big picture of the distribution of shipments across U.S. states. Logistic companies can improve their managing techniques to enhance the allocation of resources across geographic space. The natural relationship between transportation price, geography and the economy in a region that has been addressed in this study does not have empirical evidence in the current literature on this topic. The research questions have been answered by original methodologies. Pertaining to freight prices, the definition of cluster, the impact of external factors, the examination and measure of spillover effects, and testing original methods to accurately forecast truckload prices among others were formally addressed, documented and are the main contribution of this research. Finally, this work establishes a benchmark in the truckload freight transport that will aid not only logistic companies, but also

public transportation planners to understand truckload patterns in the U.S. and reliable forecasting approaches.

#### **8.4 Limitations and Recommendation for Future Research**

This section describes some limitations and indicates potential directions for further research and extensions. The availability of data and the missing information in some U.S states have limited the results. In addition data is for a third logistic company that does not involve all carriers in the U.S. Unfortunately, due to the lack of data this analysis does not provide a more specific linkage between consumer demand and TL prices except for the use of the unemployment variable. Another limitation is that the analysis uses the prices of the origin of the deliveries instead of both the origin and destination prices. Future research can be accomplished from this study, the following are recommended for future research.

1. An analysis considering both the origin and destination of the deliveries such as an interaction model between origin and destination states can be constructed to observe any advantage from previous models, and for comparing coefficient estimates. The availability of data may limit the number of states in the analysis, and the construction of the panel data structure. In addition, a regional structure determined in this study could be used for an origin-destination analysis at the regional level.
2. An exploratory and modeling analysis that considers the largest or important cities e.g. Los Angeles, New York, Houston, Tampa, Miami, etc, can be conducted. This study may be limited by the availability of data in those cities from all the years.



3. An analysis, which considers a spatial weight matrix, based on high and low demand locations of commodities for dry and reefer in the U.S. could be investigated. The exploratory analysis in this research provides information regarding those locations. Therefore, different weights can be assigned to the high, low, and medium demand locations of commodities.
4. An interesting research would be to study the connection between today's distribution of TL movements and wealth and/or income inequality in the U.S. states. Additional data sources will be required.
5. The time series analysis can be extended. A future research could check for the presence of clusters based on distances and geographic region and conduct time series data for different clusters.

## **8.5 Concluding Remarks**

Currently, big data analysis is a critical matter discussed by multiple researchers and consulting business. However, the way of dealing with data differs from research fields. In the transportation field usually data is spatially collected and this issue must be considered in the analysis. Transport and geography have an important role in the economy of any region. This research provides important information for practitioners in the trucking business and provides how to deal with large transportation data sets, in this case in the roadway freight transport. Certainly, there are many plausible approaches for forecasting prices but this research has shown how original methods perform well in the transportation field. Successful methods for forecasting TL prices are given, as well as, methods for pattern recognition. This work provides

useful insights related to TL prices, the impact of the economy in TL prices, and methods to forecast TL prices in the U.S.

## REFERENCES

- [1] American Association of State Highway and Transportation Officials (AASHTO). (2007). *America's Freight Challenge*.
- [2] ATA. American Trucking Association. (2011). *American Trucking Trends*.
- [3] Anselin, L. (1988a). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.
- [4] Anselin L. (1988b). *Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity*. *Geographical Analysis*, Vol. 20, No. 1. Ohio State University Press.
- [5] Anselin, L.; Rey, S. (1991). *Properties of Tests for Spatial Dependence in Linear Regression Models*. *Geographical Analysis*, Vol. 23, No. 2.
- [6] Anselin L. (1995). *Local Indicators of Spatial Association LISA*. *Geographical Analysis*, Vol. 27, No.2.
- [7] Anselin, L.; Bera, A.; Florax, R.; Yoon, M. (1996). *Simple Diagnostic Tests for Spatial Dependence*. *Regional Science and Urban Economics*, No 26, pp. 77-104.
- [8] Anselin, L. (1998). *Exploratory Spatial Data Analysis in a Geo-Computational Environment*. Regional Research Institute, West Virginia University.
- [9] Anselin, L.(1999a). *Interactive Techniques and Exploratory Spatial Data Analysis*. In *Geographical Information Systems. Principles, Techniques, Management and Applications*. New York: Wiley, pp. 251-264.
- [10] Anselin L. (1999b). *Spatial Econometric*. School of Social Sciences, University of Texas at Dallas.
- [11] Anselin, L.; Syabri I.; Kho Y. (2004). *GeoDa: An Introduction to Spatial Data Analysis*. Spatial Analysis Laboratory, Department of Geography, University of Illinois.
- [12] Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook*. Accessed from <https://www.geoda.uiuc.edu/documentation.php#manuals> on 15 March 2011.
- [13] Anselin, L.; Le Gallo, J.; Jayet, H. (2008). *Spatial Panel Econometrics*. *The Econometrics of Panel Data*, Chapter 19. pp. 625-659.
- [14] Baltagi, B.; Fingleton, B.; Pirotte, A. (2011). *Estimating and Forecasting with a Dynamic Spatial Panel Data Model*. Spatial Economic Research Centre.

- [15] Baumont C.; Ertur C.; Le Gallo J. (2000). *Geographic Spillover and Growth A Spatial Econometric Analysis for European Regions*. Presented at the 6th RSAI World Congress 2000 Regional Science in a Small World.
- [16] Behrens, K.; Ertur, C.; Koch, W. (2007). *Dual Gravity: Using Spatial Econometrics to Control for Multilateral Resistance*. Center for Operations Research and Econometrics (CORE), Université Catholique de Louvain.
- [17] Beilock, R; Garrod, P; Millius, W. (1986). *Freight Charge Variations in Truck Transport Markets: Price Discrimination or Competitive Pricing?*. American Journal of Agricultural Economics, Vol. 68, No. 2, pp. 226-236. Oxford University Press.
- [18] Bera, A.; Yoon, M. (1993). *Specification Testing with Locally Misspecified Alternatives*. Econometric Theory. Cambridge University Press. Vol. 9, No. 4, pp. 649-658.
- [19] Besag, J. (1981). *On Resistant Techniques and Statistical Analysis*. Biometrika, 68, 2, 463-9
- [20] Bolker, B. (2008). *Exploratory Data Analysis and Graphics*. Ecological Models and Data in R. Chapter 2. Accessed from [http://press.princeton.edu/chapters/s02\\_8709.pdf](http://press.princeton.edu/chapters/s02_8709.pdf) on 05 February 2011.
- [21] Boyandin, I.; Bertini, E.; Lalanne, D. (2010). *Using Flow Maps to Explore Migrations Over Time*. In: Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science (GeoVA).
- [22] Boyandin, I.; Bertini, E.; Bak, P.; Lalanne, D. (2011). Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data. Eurographics/IEEE-VGTC Symposium on Visualization, Computer Graphics Forum, International Journal of the Eurographics Association.
- [23] BLS. U.S. Bureau of labor Statistics. (2012). *The Recession of 2007–2009*. Spotlight on Statistics. Accessed from [www.bls.gov/spotlight](http://www.bls.gov/spotlight) on 10 April 2012.
- [24] Bureau of Economic Analysis. United States Department of Commerce. Accessed from [www.bea.gov](http://www.bea.gov) on 20 May 2011.
- [25] Bureau of Labor Statistics. United States Department of Labor. Accessed from [www.bls.gov](http://www.bls.gov) on 25 May 2011.
- [26] Cameron, A.; Trivedi, P. (2005). *Microeconomics: Methods and Applications*. Linear Panel Models: Basic. Cambridge U. Press. Chapter 21.
- [27] Capello, R. (2009). *Spatial Spillovers and Regional Growth: A Cognitive Approach*. European Planning Studies. Vol 7, No 5.

- [28] Chatfield, C. (1986). *Exploratory Data Analysis*. European Journal of Operational Research, 23, 5-13.
- [29] Cho, T.; Newman D. (2005). *Spatial Analysis of Rural Land Development*. Journal of Forest Policy and Economics.
- [30] Church, J.; Ware R. (1999). *Industrial Organization: A Strategic Approach*. New York: McGraw-Hill. Available on: [http://works.bepress.com/jeffrey\\_church/23](http://works.bepress.com/jeffrey_church/23)
- [31] Clark, T.; Linzer D. (2012). *Should I Use Fixed or Random Effects?*. Emory University.
- [32] Cliff, A., Ord, K. (1981). *Spatial Processes: Models and Applications*. London: Pion.
- [33] Cox, N; Jones, K. (1981). *Exploratory Data Analysis*. Wrigley, N. and Bennett, R.J.,eds. Quantitative Geography: a British view, Routledge and Kegan Paul. Chapter 13.
- [34] Crandall M.; Weber, B. (2004). *Local Social and Economic Conditions, Spatial Concentrations of Poverty, and Poverty Dynamics*. Rural Poverty Research Center. RUPRI.
- [35] Debarsy, N.; Ertur, C.; LeSage, J. (2011). *Interpreting Dynamic Space–Time Panel Data Models*. Statistical Methodology.
- [36] Dougherty, C. (2007). *Introduction to Econometrics*. Thrid Edition. Oxford University Press. Chapter 14.
- [37] Dubin, R. (1998). *Spatial Autocorrelation: A Primer*. Journal of Housing Economics, No 7, pp. 304–327.
- [38] Durango, P. (2006). *A Time Series Analysis Framework for Transportation Infrastructure Management*. Science Direct. Transportation Research Part B.
- [39] Elhorst, P. (2011). *Dynamic Spatial Panels: Models, Methods, and Inferences*. Geographic Systems. Springer-Verlag.
- [40] Elhorst, P. (2010). *MATLAB Software for Spatial Panel*. Presented at the IVth World Conference of the Spatial Econometrics Association (SEA), Chicago, June 9-12.
- [41] EIA. U.S. Energy Information Administration. Accesed from <http://www.eia.gov> on 12 September 2010.
- [42] Florax, R.; Folmer, H.; Rey, S. (2003). *Specification Searches in Spatial Econometrics: The Relevance of Hendry’s Methodology*. Regional Science and Urban Economics, Vol 33, No 5, pp. 557–579.

- [43] Flowmap Software. (2010). Human Geography and Planning Department. Utrecht University. Accessed from <http://flowmap.geog.uu.nl/index.php> on 15 February 2012.
- [44] FlowMapping. (2010). Spatial Data Mining and Visual Analytics Lab. Geographic Knowledge Discovery with Human-Computer Collaboration. Department of Geography University of South Carolina. Accessed from <http://www.spatialdatamining.org/> on 10 March 2012.
- [45] FHWA. Federal Highway Administration. (2011). *Freight Facts and Figures*. Office of Freight Management and Operation.
- [46] Gahegan, M. (1999). *Four Barriers to the Development of Effective Exploratory Visualization Tools for the Geosciences*. International Journal of Geographical Information Systems, No 13, pp. 289-309.
- [47] GeoDa. Center for Geospatial Analysis and Computation. School of Geographical Science and Urban Planning. Accessed from <https://geodacenter.asu.edu/node/402#choice> on 25 January 2012.
- [48] Good, I. (1983). *The Philosophy of Exploratory Data Analysis*. Philosophy of Science, No 50, pp. 283-295.
- [49] Groff, E.; McEwen, T. (2006). *Visualization of Spatial Relationships in Mobility Research: A Primer*. Prepared for National Institute of Justice.
- [50] Handbook of Statistical Methods. (2006). NIST/SEMATECH. Accessed from <http://www.itl.nist.gov/div898/handbook/> on 05 June 2011.
- [51] Haining, R. (2009). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press. Chapters 5, 6, 7.
- [52] Haire, A.; Machemehl, R. (2009). *A Methodology for Incorporating Fuel Price Impacts into Short-term Transit Ridership Forecasts*. Southwest Region University Transportation Center.
- [53] Hall, R. (1985). *Dependence Between Shipment Size and Mode in Freight*. Transportation Science. Vol. 19, No. 4, pp. 436-444.
- [54] Hartwig, F.; Dearing, B. (1979). *Exploratory Data Analysis*. Quantitative Applications in the Social Sciences, No 16.
- [55] Hu, A.; Liu, S. (2010). *Transportation, Economic Growth and Spillover Effects: The Conclusion Based on the Spatial Econometric Model*. Higher Education Press. Springer-Verlag.

- [56] Jflowmap. (2012). University of Fribourg. Accessed from <http://code.google.com/p/jflowmap/> on 10 July 2012.
- [57] Kay, M.; Warsing, D. (2009). *Estimating LTL Rates Using Publicly Available Empirical Data*. International Journal of Logistics: Research and Applications. Vol. 12, No. 3, pp. 165-193.
- [58] Krugman, P.R. (1991). *Geography and Trade*. Gaston Eyskens Lecture Series. MIT Press, Cambridge, MA. pp. 153.
- [59] Lacombe, D.; Holloway, G.; Shaughnessy, T. (2012). *Bayesian Estimation of the Spatial Durbin Error Model with an Application to Voter Turnout in the 2004 Presidential Election*. International Regional Science Review. SAGE .
- [60] Lee, S. (2009). *Bias from Misspecified Spatial Weight Matrices in SAR Models: Theory and Simulation Studies*. Department of Economics, San Francisco State University.
- [61] LeSage, J. (1997). *Regression Analysis of Spatial Data*. Journal of Regional Analysis and Policy.
- [62] LeSage, J. (1999). *The Theory and Practice of Spatial Econometrics*. Department of Economics. University of Toledo.
- [63] LeSage, J.; Pace, K. (2005). *Spatial Econometric Modeling of Origin-Destination Flows*. Department of Economics, University of Toledo.
- [64] LeSage, J.; Polasek, W. (2006). *Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows*. Institute for Advanced Studies, Vienna.
- [65] LeSage, J; Pace, K. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs.
- [66] Lewis, C. D. (1982). *Industrial and Business Forecasting Method*. Butter Worth Scientific, London.
- [67] Mayor, M.; Patuelli, R. (2012). *Short-Run Regional Forecasts: Spatial Models through Varying Cross-Sectional and Temporal Dimensions*. The Rimini Centre for Economic Analysis.
- [68] Miller, J. (1997). *Spatial Aggregation and Regional Economic Forecasting*. The Analysis of Regional Science. Springer-Verlag.
- [69] Moran, P. (1950). *Notes on Continuous Stochastic Phenomena*. Biometrika, No 37, pp. 17.

- [70] Novak, D.; Hodgdon, C.; Feng, G.; Aultman-Hall, L. (2008). *Nationwide Freight Generation Models: A Spatial Regression Approach*. Springer Science.
- [71] Özkaya, E.; Keskinocak, P.; Roshan, J.; Weight, R. (2010). *Estimating and Benchmarking Less-Than-Truckload Market Rates*. Transportation Research Part E, Vol 46, pp. 667-682.
- [72] Pace, K.; Barry, R.; Sirmans, C. (1998). *Spatial Statistics and Real Estate*. Journal of Real Estate Finance and Economics, Vol 17, No 1, pp. 5-13.
- [73] Paelinck, J.; Klaassen, L. (1979). *Spatial Econometrics*. Saxon House Farnborough.
- [74] Parent, O.; LeSage, J. (2010). *A Spatial Dynamic Panel Model with Random Effects Applied to Commuting Times*. Transportation Research Part B.
- [75] Pennings, P.; Keman, H.; Kleinnijenhuis, J. (1999). *Doing Research in Political Science: An Introduction to Comparative Methods and Statistics*. SAGE.
- [76] Pitfield, D. (2004). *Airline Price Competition: A Time Series Analysis of Low-Cost Carriers*. Transport Studies Group, Loughborough University. Presented to the 44th European Congress of the Regional Science Association International.
- [77] Podesta, F. (2000). *Recent Developments in Quantitative Comparative Methodology: The case of Pooled Time Series Cross-Section Analysis*. McDonough School of Business, Georgetown University.
- [78] Pollock, S. (1993). *A Short Course of Time Series Analysis and Forecasting*. Queen Mary and West eld College, the University of London.
- [79] Quddus, M. (2008). *Modeling Area-Wide Count Outcomes with Spatial Correlation and Heterogeneity: An Analysis of London Crash Data*. Journal of Accident Analysis and Prevention.
- [80] Rey, S.; Montouri, B. (1999). *U.S. Regional Income Convergence: A Spatial Econometric Perspective*. Regional Studies Association.
- [81] RITA. Research and Innovative Technology Administration. (2009). *Estimation and Analysis of Multifactor Productivity in Truck. Transportation:1987-2003*. Bureau of Transportation Statistics
- [82] Rodrigue, J-P; Comtois, C.; Slack, B. (2009), *The Geography of Transport Systems*, Second Edition, New York: Routledge.
- [83] Saffet, E. (2009). *Explorative Spatial Analysis of Traffic Accident Statistics and Road Mortality Among the Provinces of Turkey*. Journal of Safety Research.

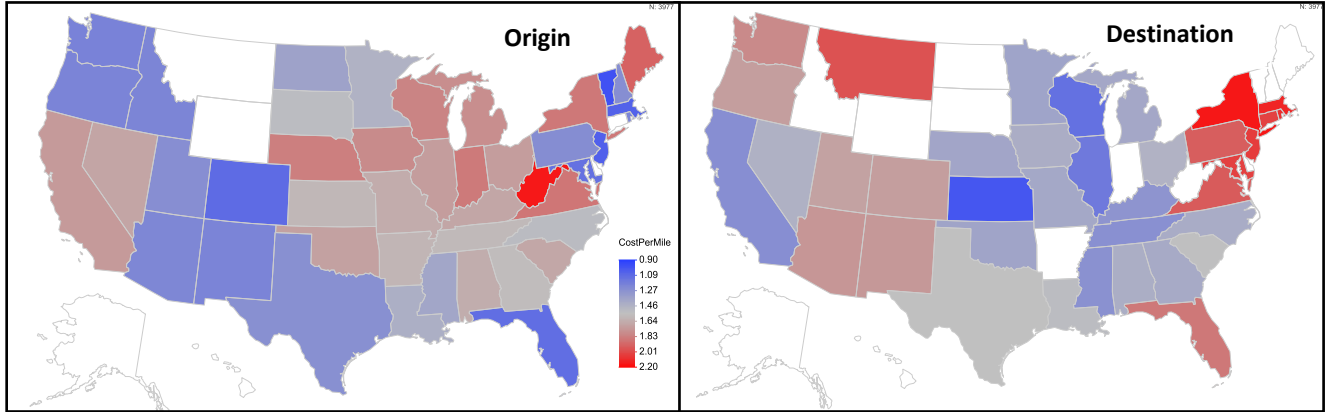


- [84] Söderbom, M. (2011). *Econometric II: Panel Data Part I*. Department of Economics, University of Gothenburg. Accessed from [www.economics.gu.se/soderbom](http://www.economics.gu.se/soderbom) on 02 February 2012.
- [85] Swenseth, S.; Godfrey, M. (2001). *Incorporating Transportation Costs Into Inventory Replenishment Decisions*. International Journal of Production Economics. No. 77, pp 113- 130.
- [86] Swenseth, S.; Godfrey, M. (1996). *Estimating Freight Rates for Logistics Decisions*. Journal of Business Logistics, Vol 17, No 1, pp. 213-231.
- [87] SelectUSA. (2012). *The Logistics and Transportation Industry in the United States Logistics and Transportation*. The U.S. Logistics and Transportation Industry. Accessed from <http://selectusa.commerce.gov> on 18 August 2011.
- [88] Thaiprasert, N.; Hicks, M. (2011). *The Effect of Higher Fuel Prices on Indiana's Economy*. Center for Business and Economic Research. Ball State University.
- [89] Tobler, W. (1987). *An Experiments in Migration Mapping by Computer*. The American Cartographer, No 14, pp. 155-163
- [90] Tobler W., (1970). *A Computer Movie Simulating Urban Growth in the Detroit Region*. Economic Geography, No 46, pp. 234-240.
- [91] Tukey, J. (1977). *Exploratory Data Analysis*. Addison Wesley, Reading, MA.
- [92] Tukey, J.; Wilk, M. (1970). *Data Analysis and Statistics: Techniques and Approaches*. In Tufte, E. R. (ed) *The Quantitative Analysis of Social Problems*, 370-390, Addison-Wesley, Reading, Mass.
- [93] Velleman, P.; Hoaglin, D. (1981). *The ABC's of EDA: Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury.
- [94] Wells, J. (2009). *The Importance of Transportation Forecasting*. U.S. Department of Transportation. Workshop for Transportation Forecasters.
- [95] Williams, M.; Hoel, L. (2003). *Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results*. Journal of Transportation Engineering. ASCE.
- [96] Yang, Y.; Wong, K. (2012). *A Spatial Econometric Approach to Model Spillover Effects in Tourism Flows*. Journal of Travel Research.
- [97] Yong, K. (2006). *Spatial Econometric Analysis of Highways and Regional Economy in Missouri*. Dissertation. Department of Economics, University of Missouri-Columbia.

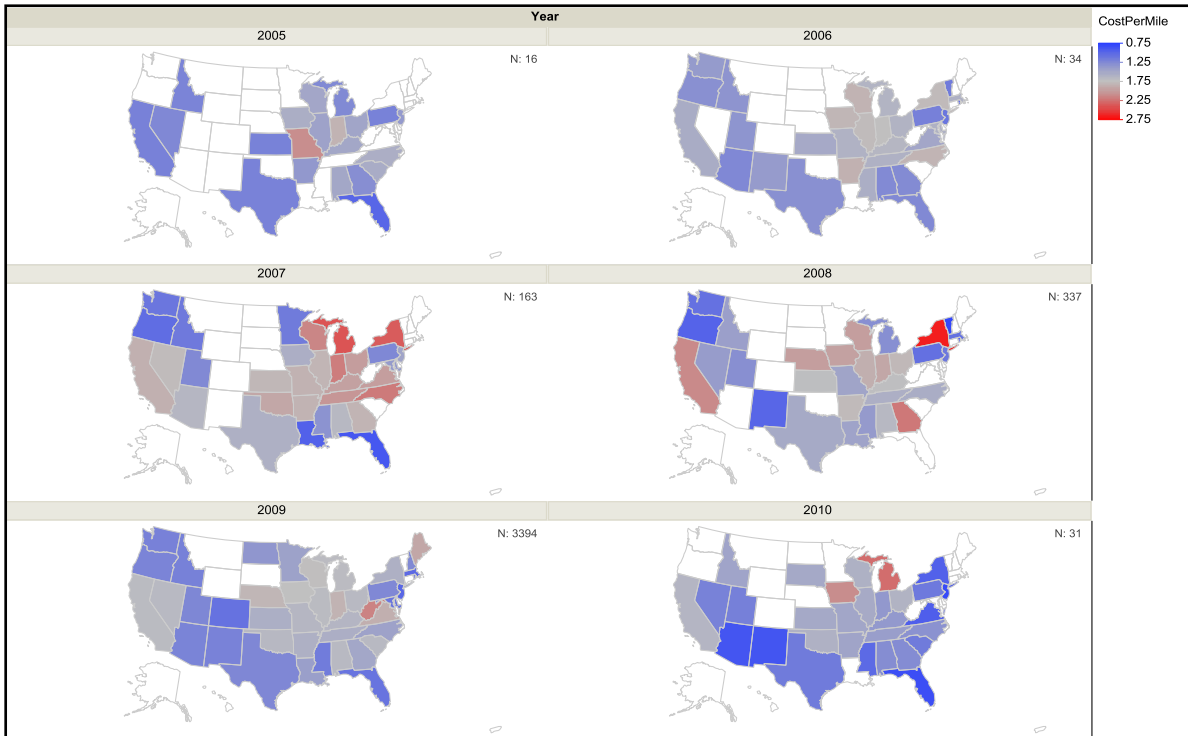
- [98] Yu, J.; Jong, R.; Lee, L. (2011). *Estimation for Spatial Dynamic Panel Data with Fixed Effects: The Case of Spatial Cointegration*. Third Symposium on Econometric Theory and Applications at HKUST.
- [99] Zhenhua, C.; Kingsley, E. (2012). *Surface Transportation and Regional Output: A Spatial Panel Approach*. George Mason University. School of Public Policy. Research Paper No. 2012-15.

## APPENDIX A: Exploratory Maps

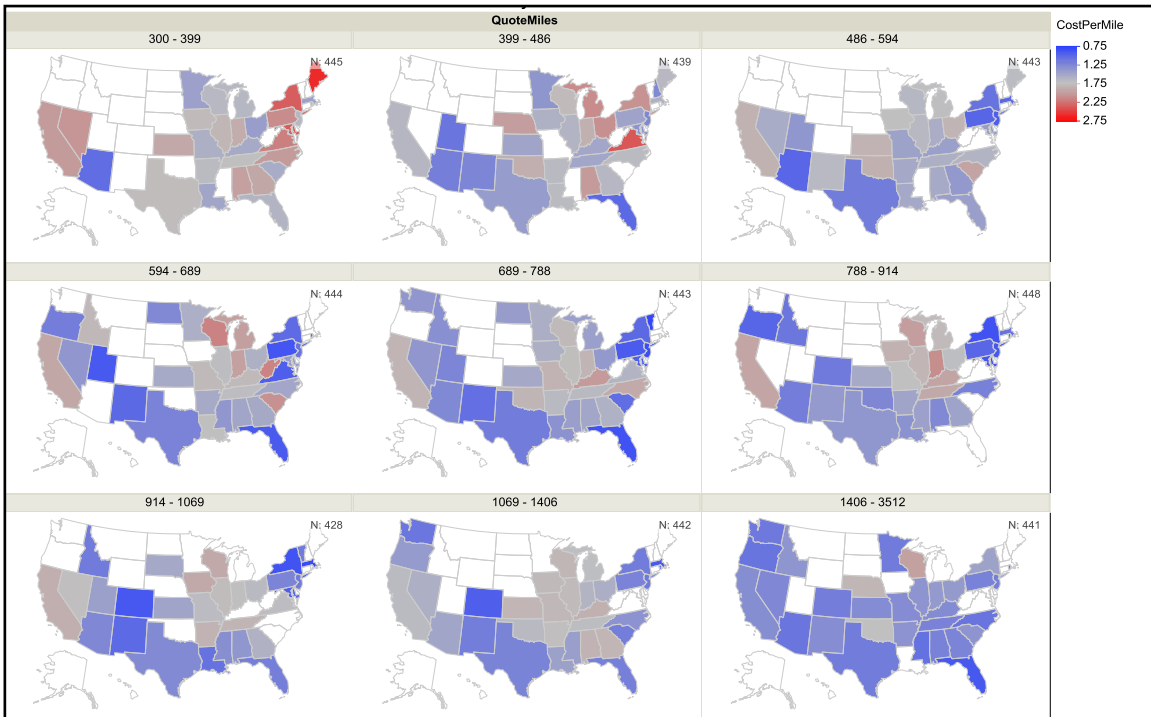
### A.1. Dry Load Managed Lanes



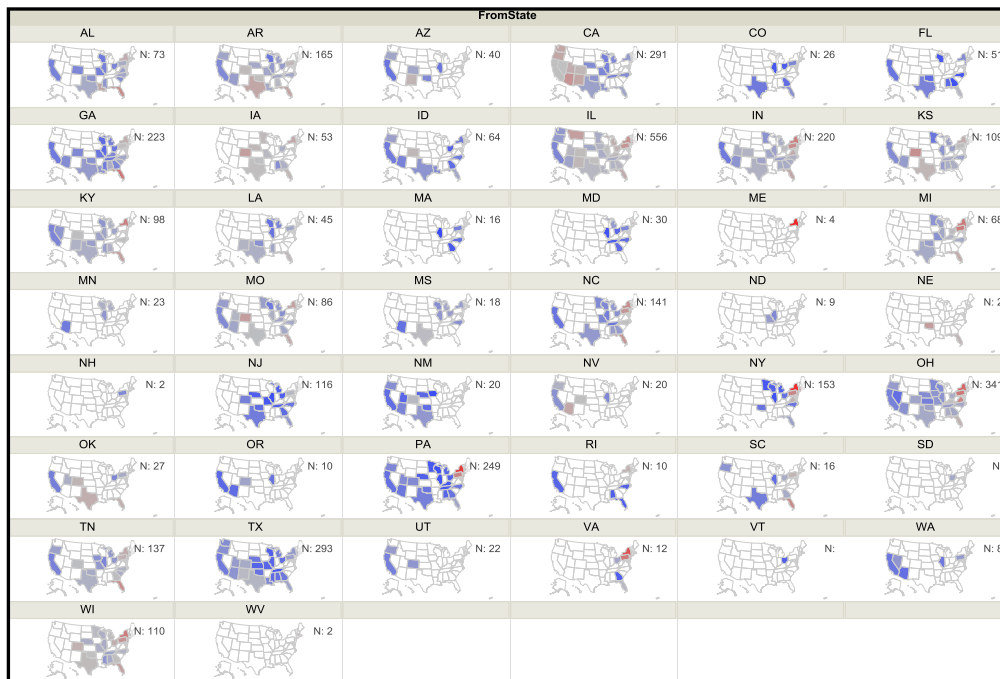
**Figure A.1.1 Dry Load Managed - Six Years Averaged**



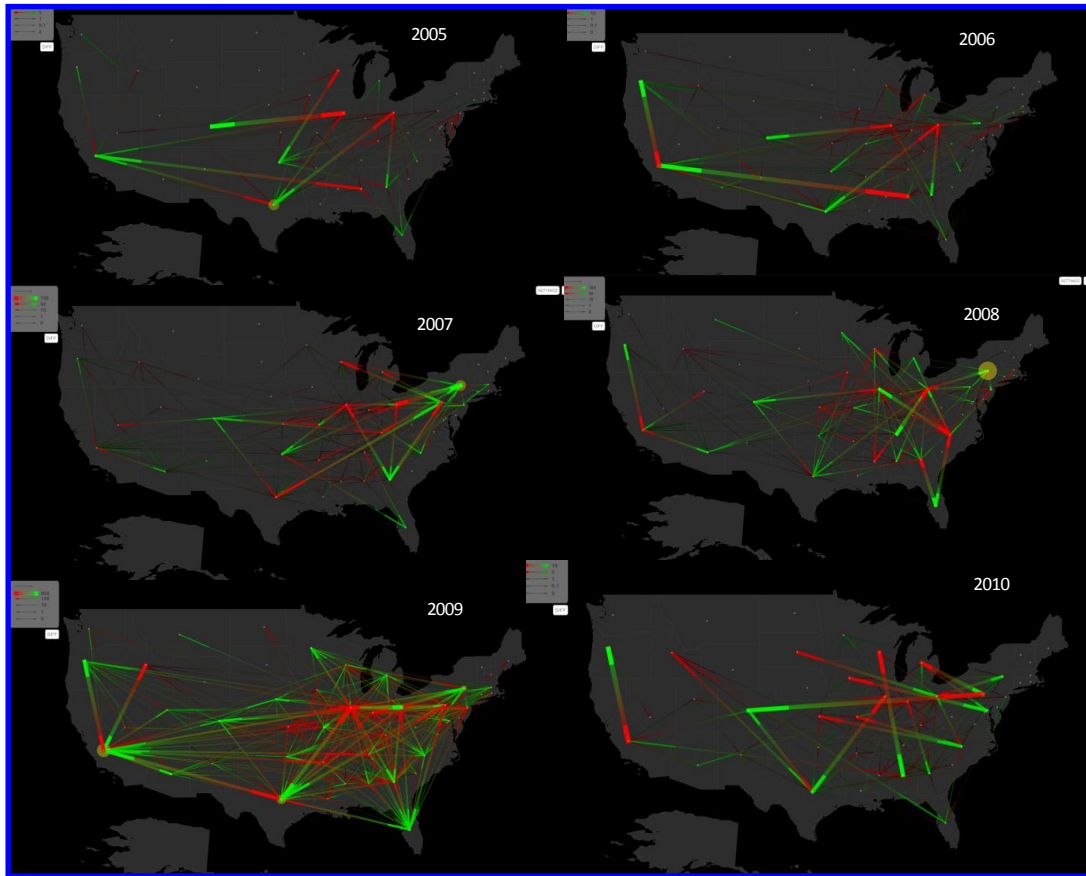
**Figure A.1.2: Dry Load Managed by Year**



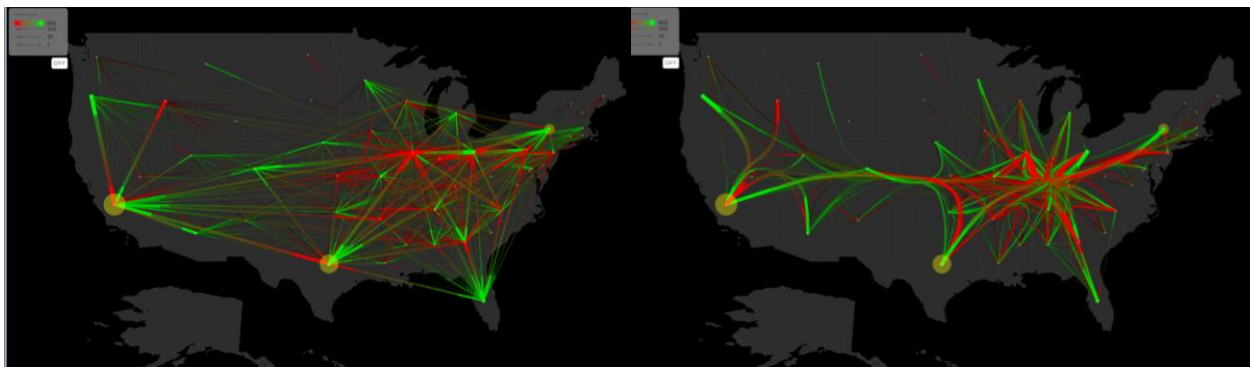
**Figure A.1.3: Averaged TL price for Dry Load Managed by Distance**



**Figure A.1.4: Average TL Price for Dry Load Managed by O-D State**

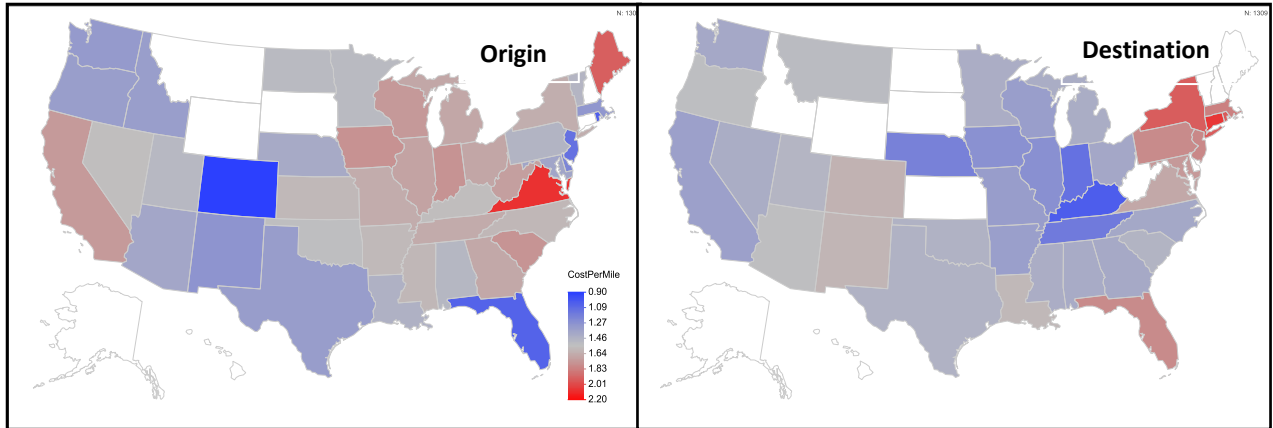


**Figure A.1.5: Flow Map over the Years: Dry Load Managed**

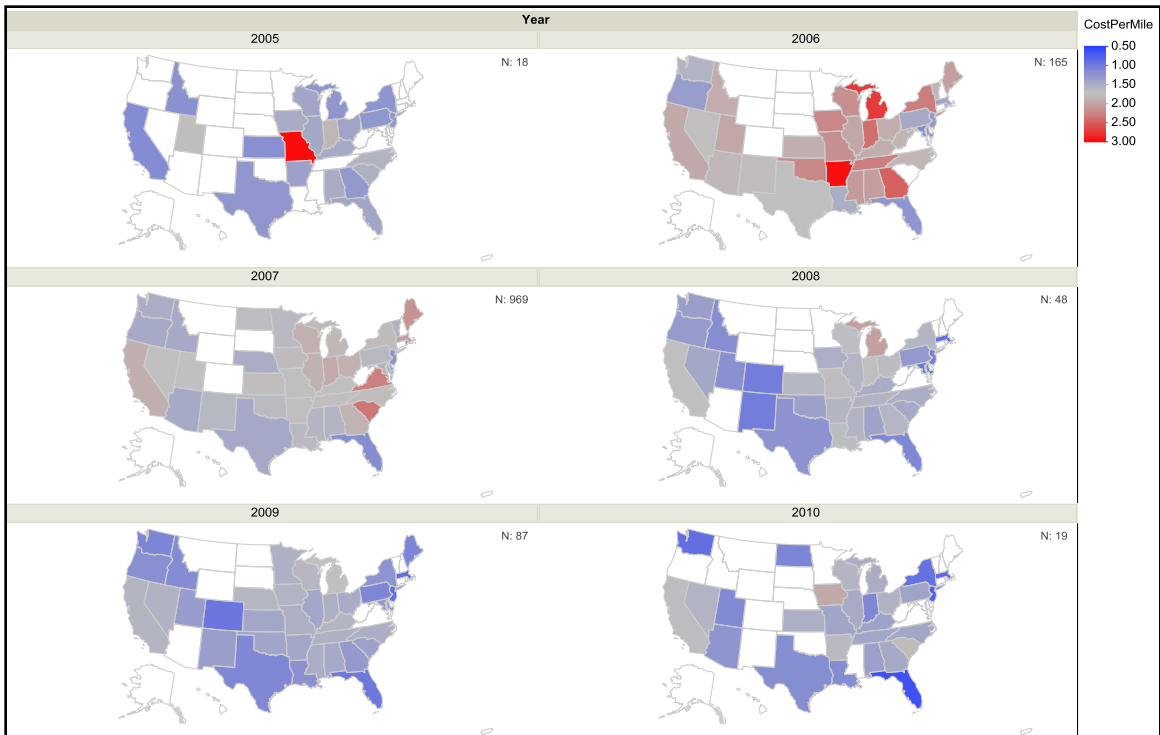


**Figure A.1.6: Flow Map - Total: Dry Load Managed**

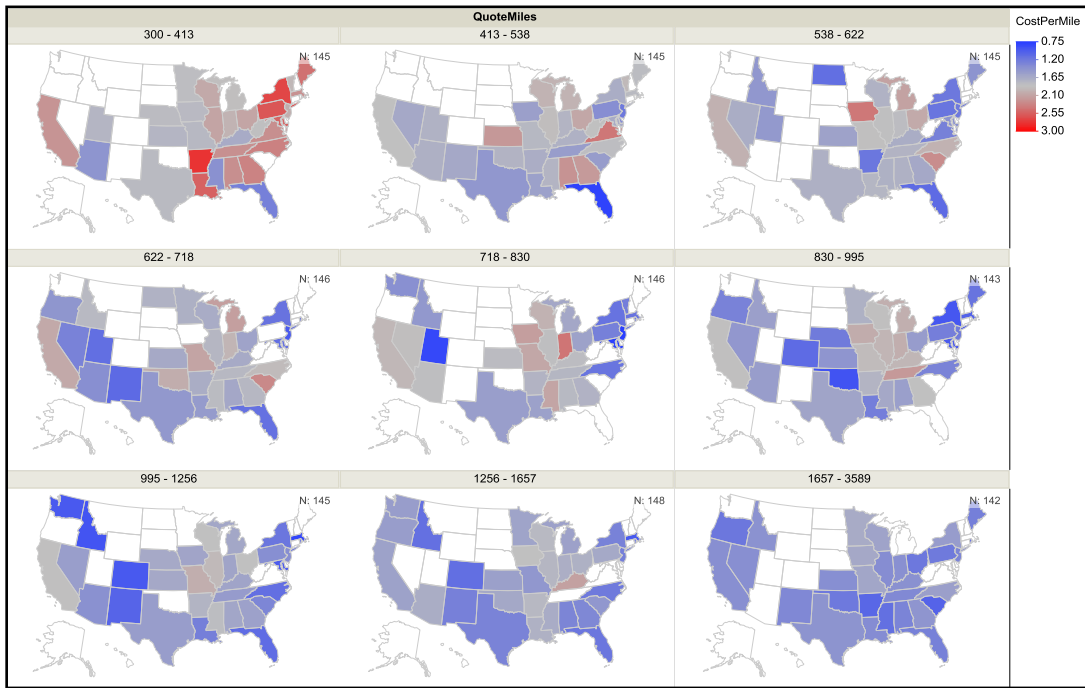
## A.2 Dry Loads non-Managed Lanes



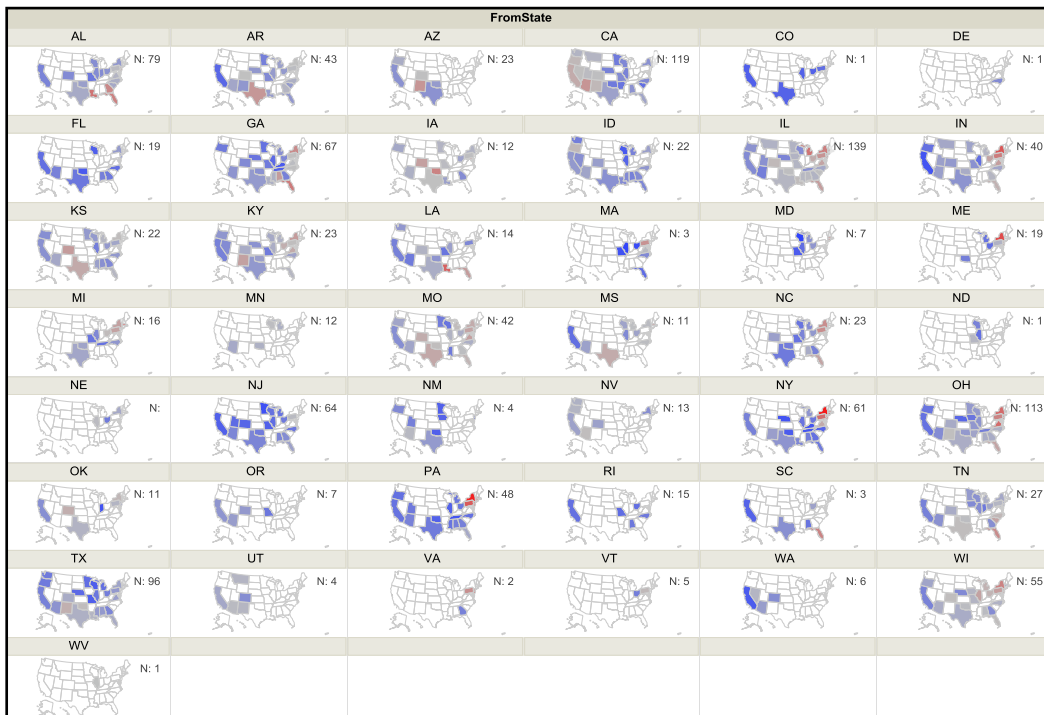
**Figure A.2.1: Dry Load Non-Managed - Six Years Averaged**



**Figure A.2.2: Dry Load Non-Managed by Year**

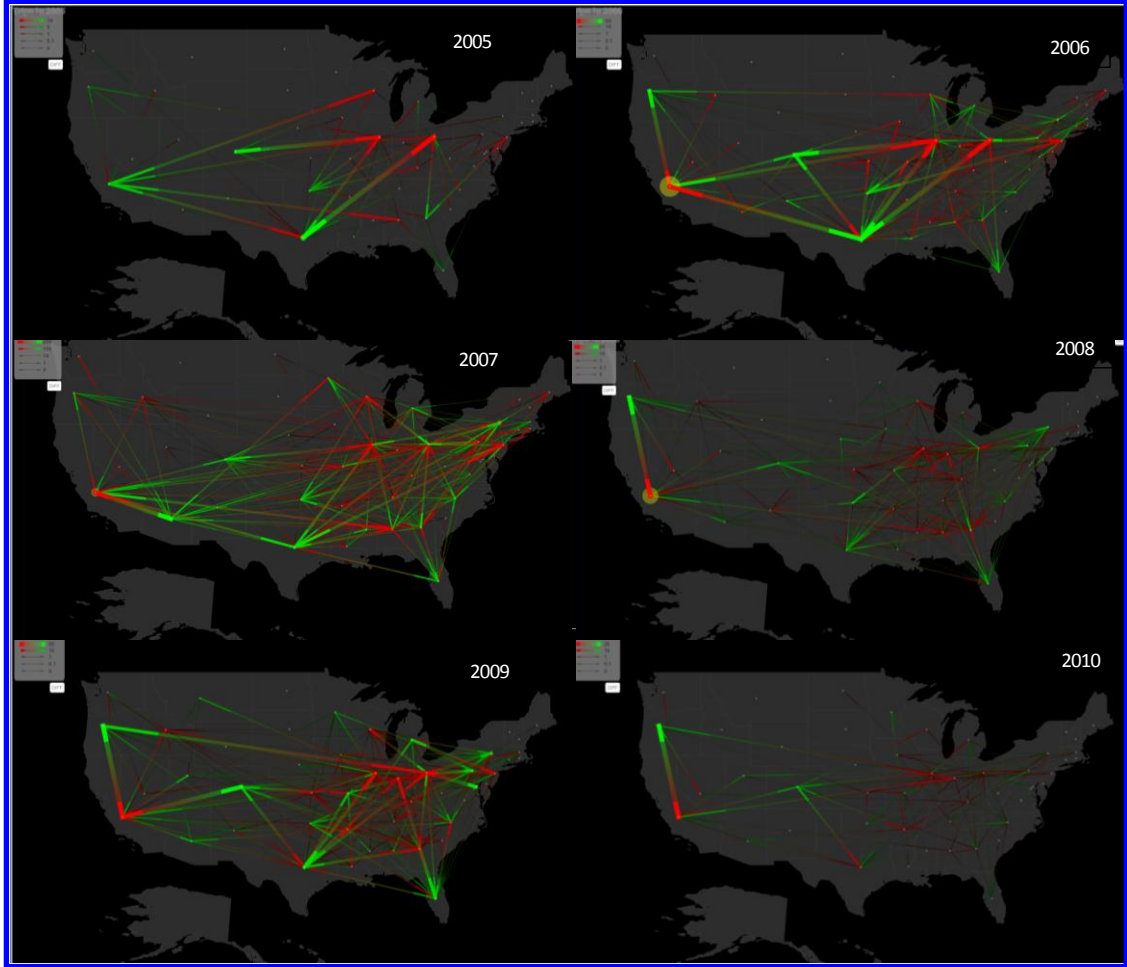


**Figure A.2.3: Averaged TL price for Dry Load Non-Managed by Distance**

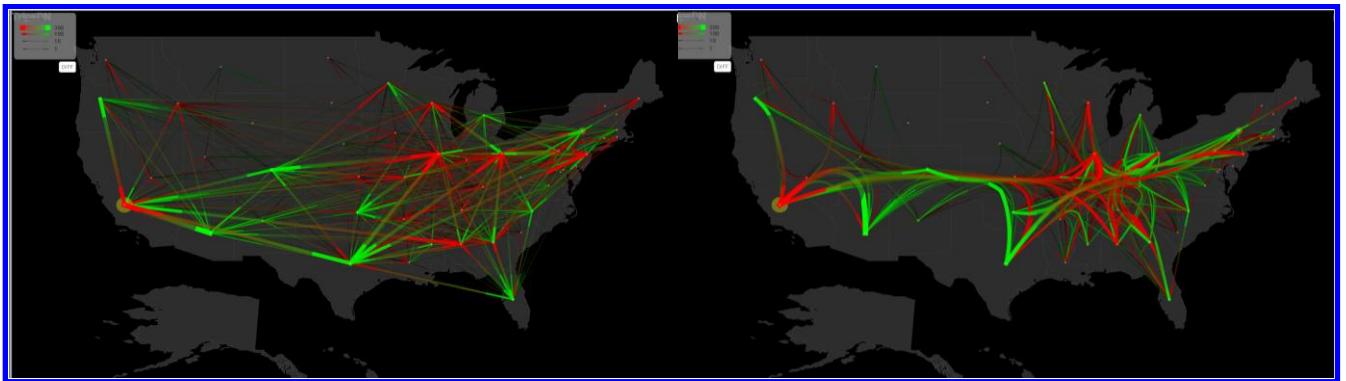


**Figure A.2.4: Average TL Price for Dry Load Non-Managed by O-D State**





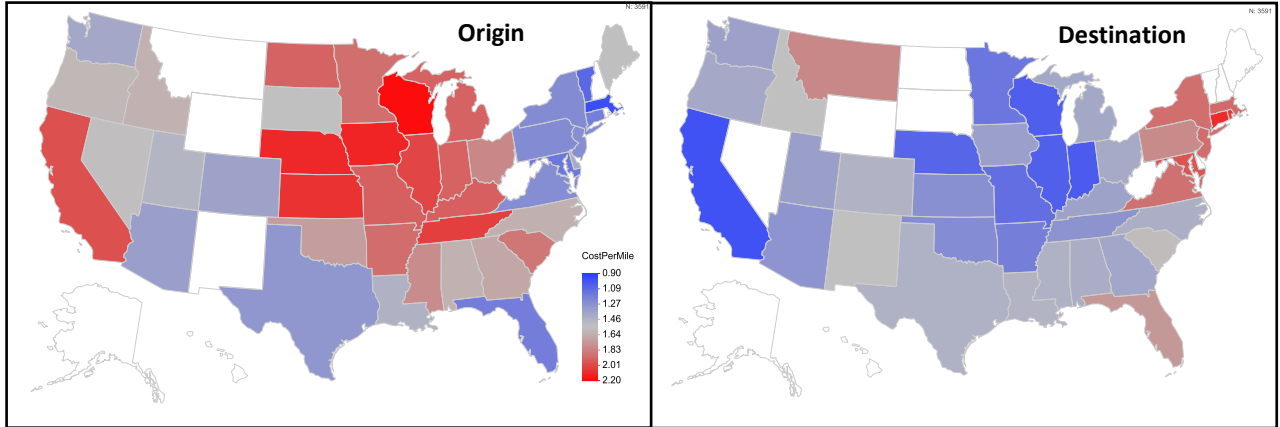
**Figure A.2.5: Flow Map over Years: Dry Load non-Managed**



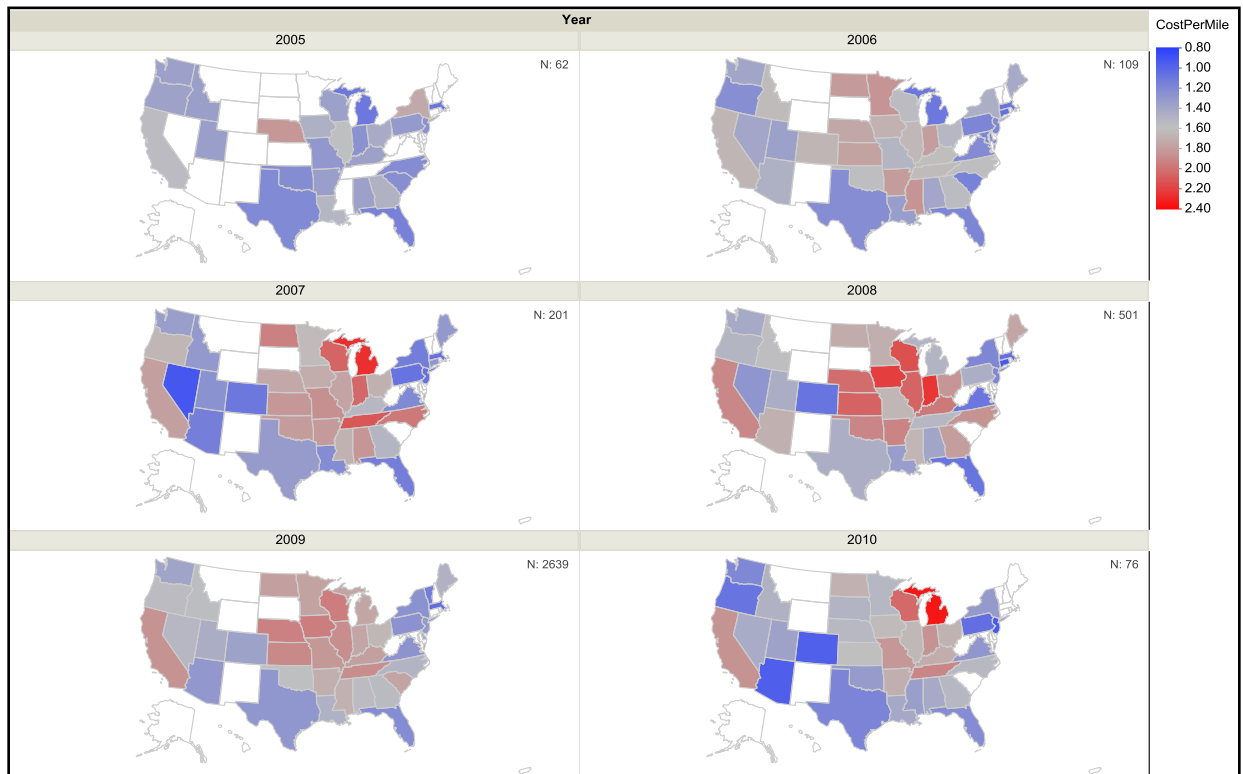
**Figure A.2.6: Flow Map - Total: Dry Load non-Managed**



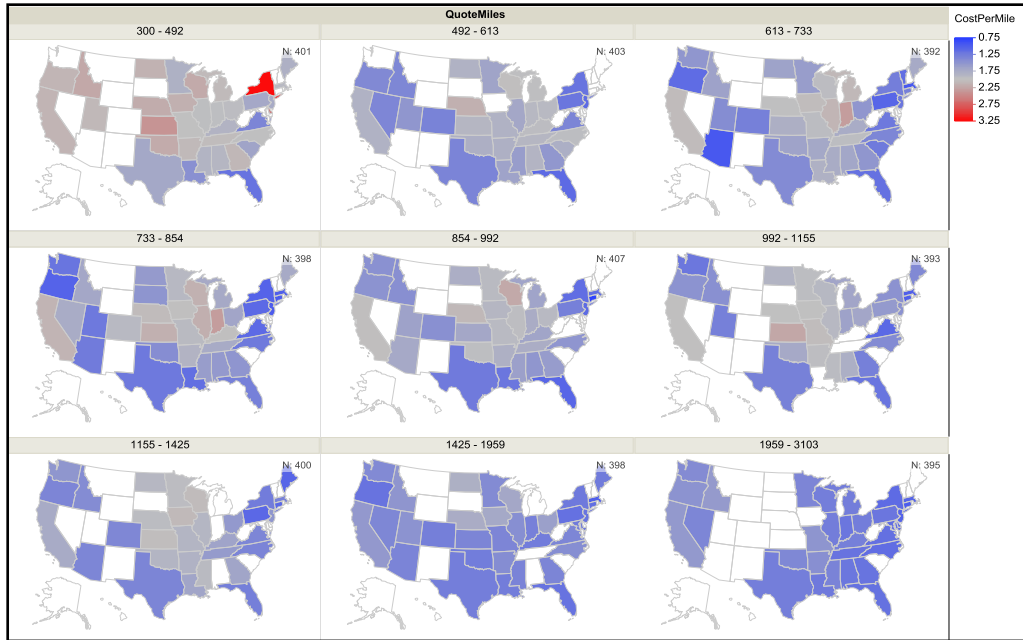
### A.3 Reefer Loads - Lane Managed



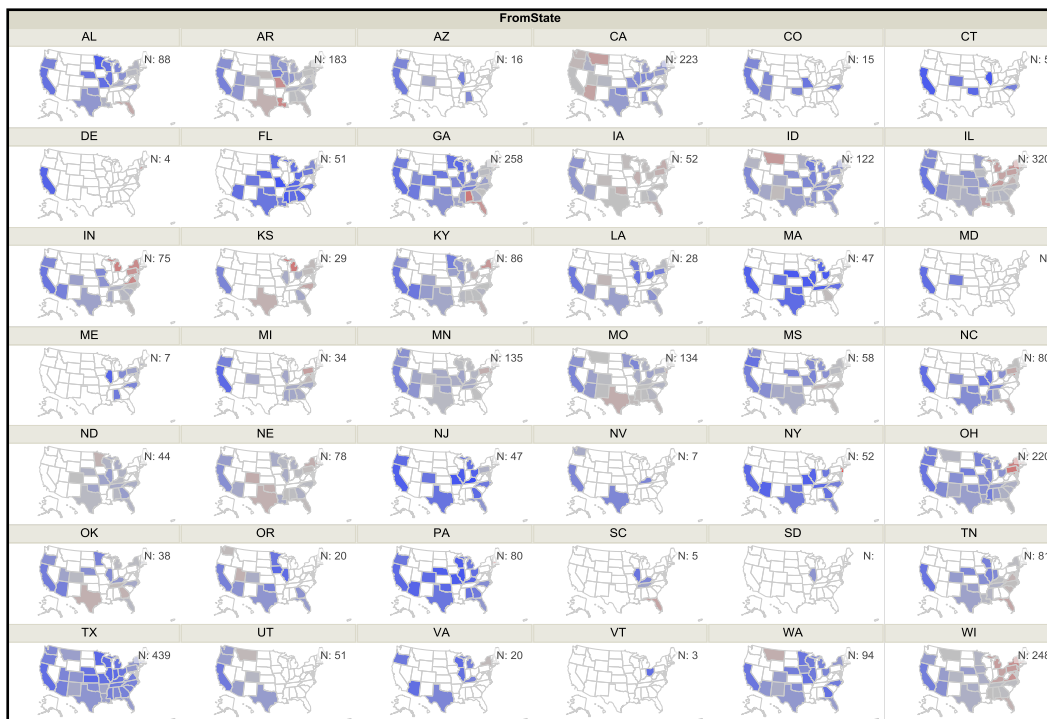
**Figure A.3.1: Reefer Load Managed - Six Years Averaged**



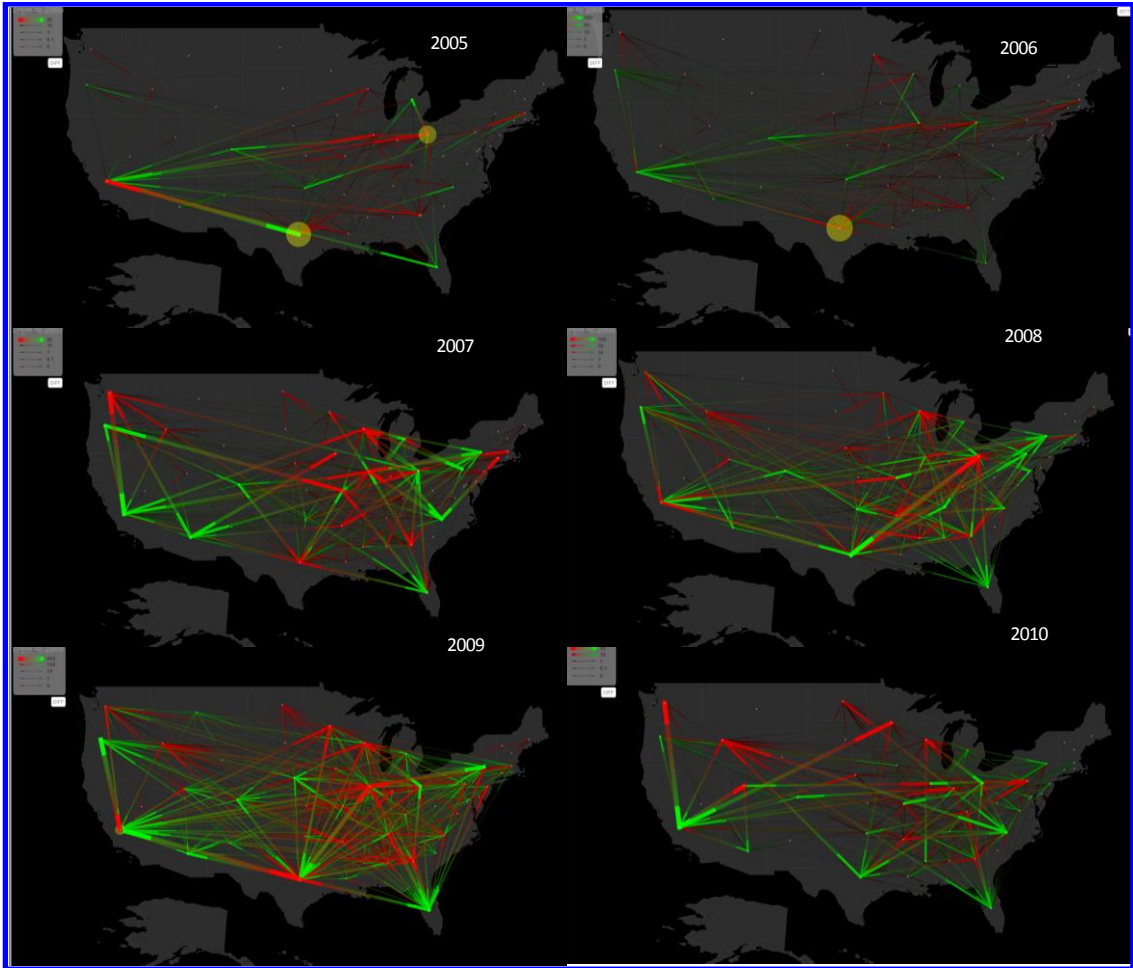
**Figure A.3.2: Reefer Load Managed by Year**



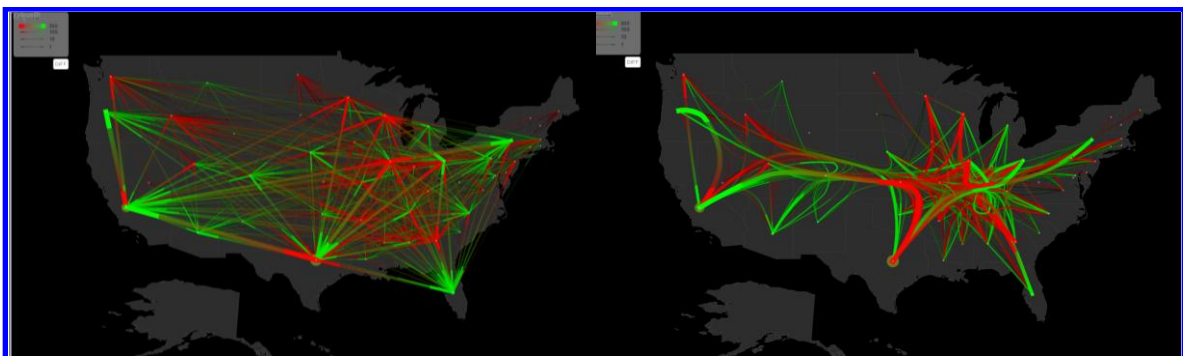
**Figure A.3.3: Averaged TL price for Reefer Load Managed by Distance**



**Figure A.3.4: Average TL Price for Reefer Load Managed by O-D State**

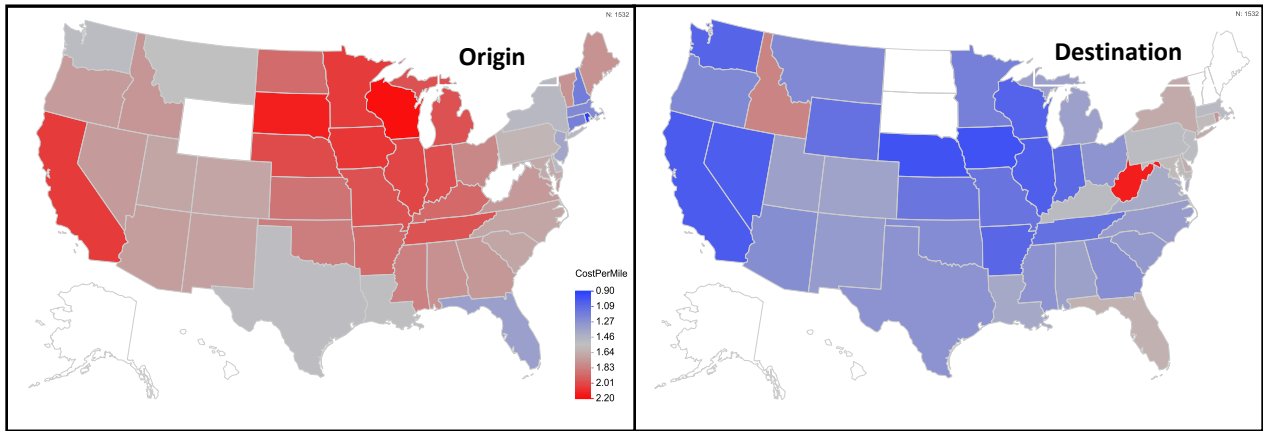


**Figure A.3.5: Flow Map over Years: Reefer Load Managed**

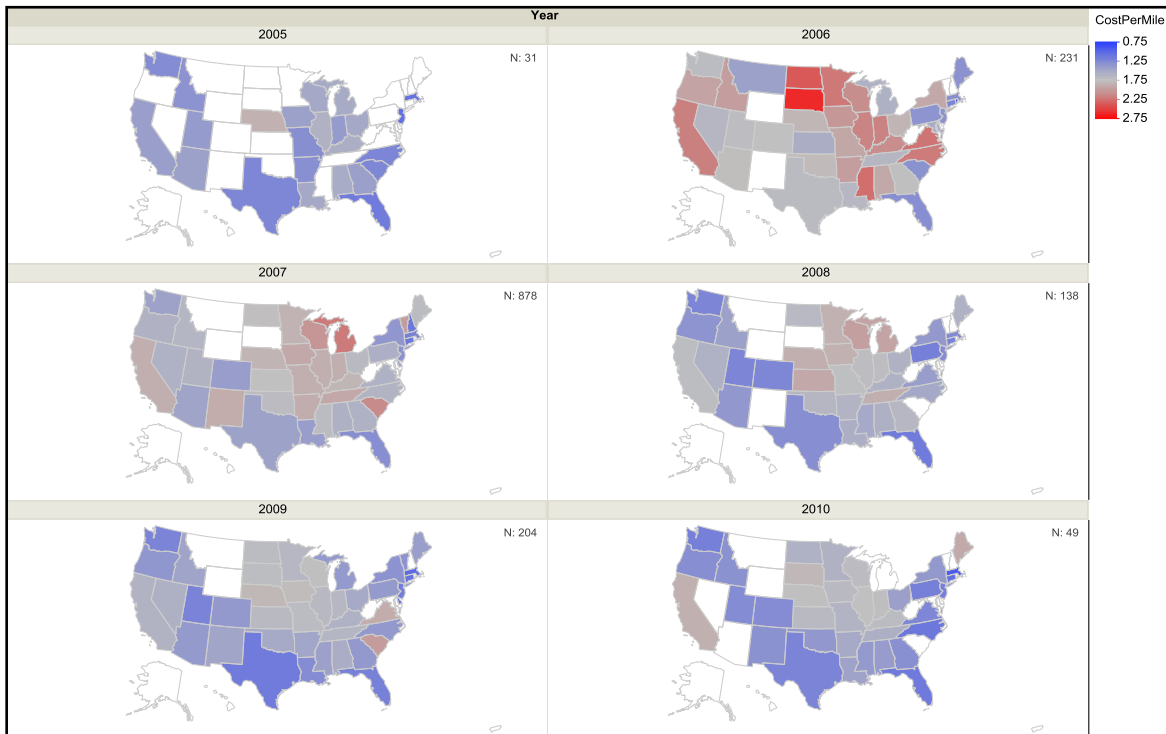


**Figure A.3.6: Flow Map - Total: Reefer Load Managed**

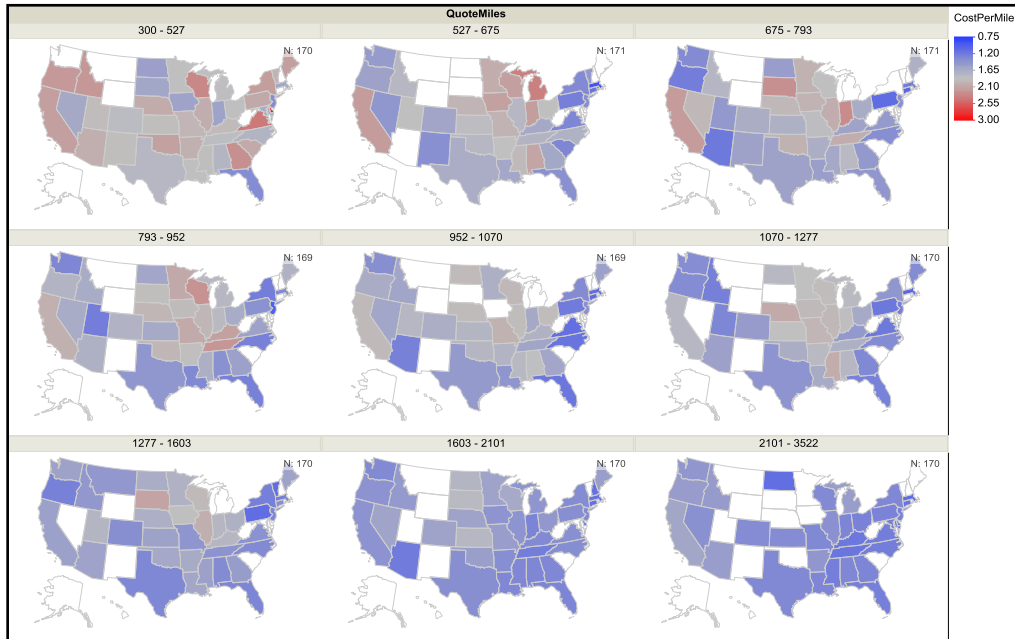
## A.4. Reefer Load non-Managed Lanes



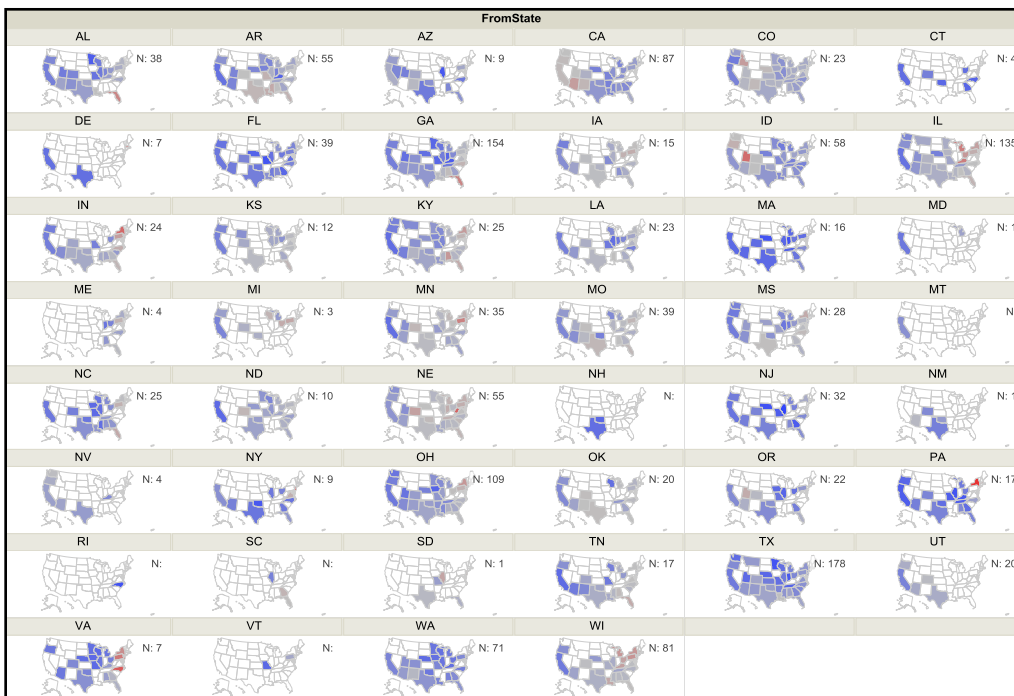
**Figure A.4.1: Reefer Load Non-Managed - Six Years Averaged**



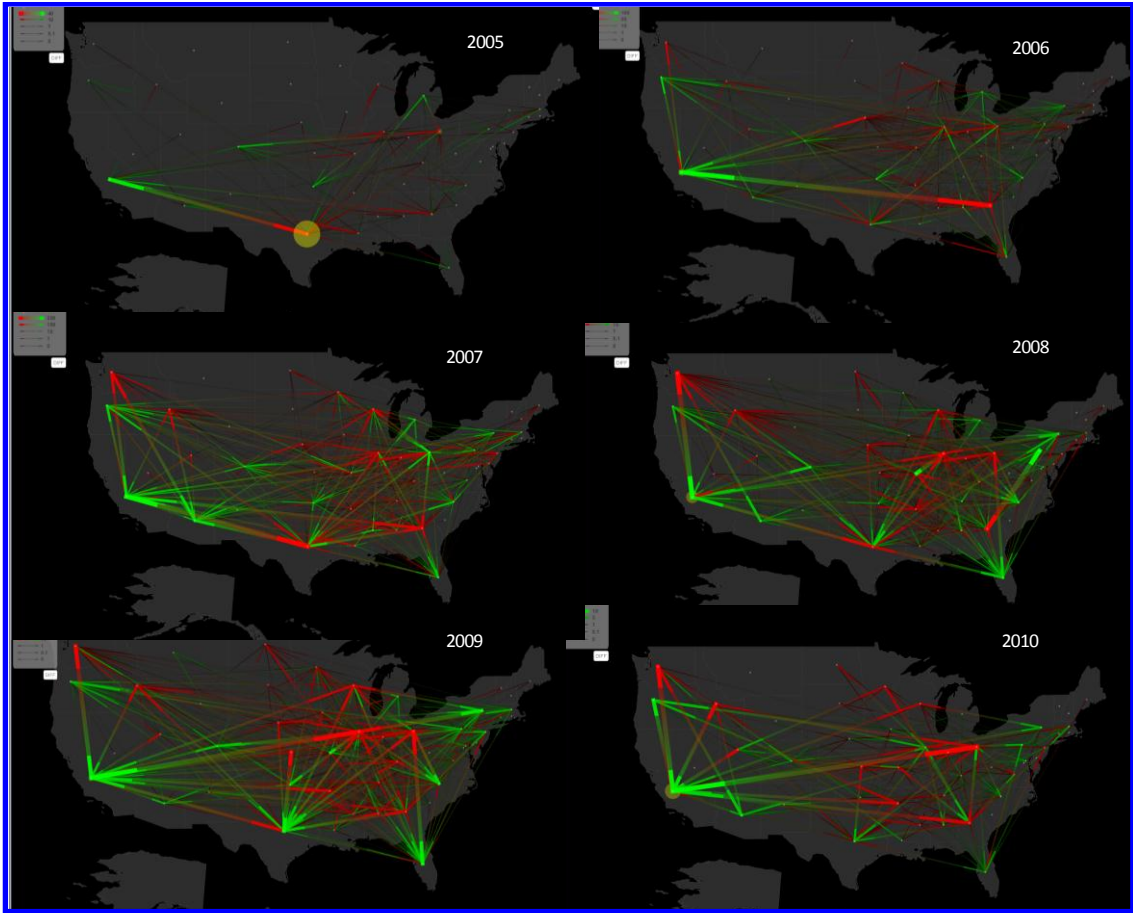
**Figure A.4.2: Reefer Load Non-Managed by Year**



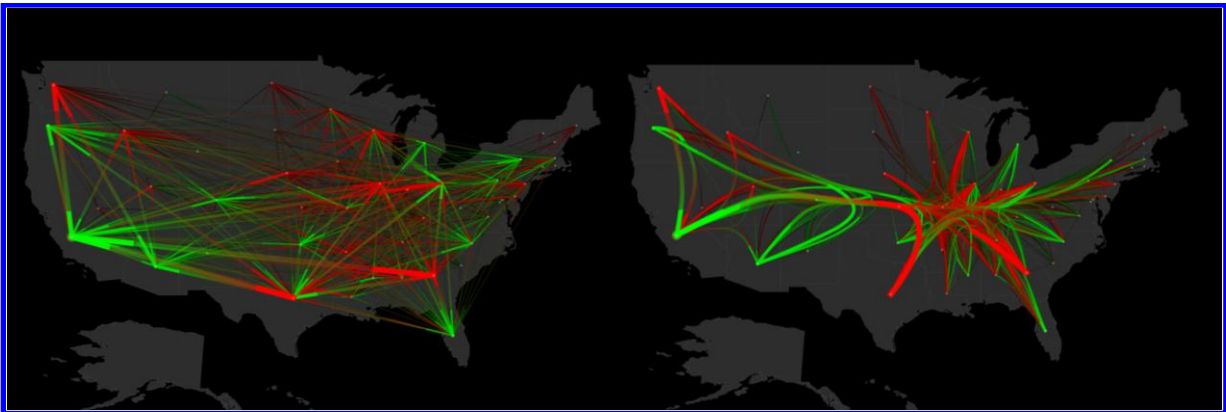
**Figure A.4.3: Averaged TL price for Reefer Load Non-Managed by Distance**



**Figure A.4.4 Average TL Price for Reefer Load Non-Managed by O-D State**



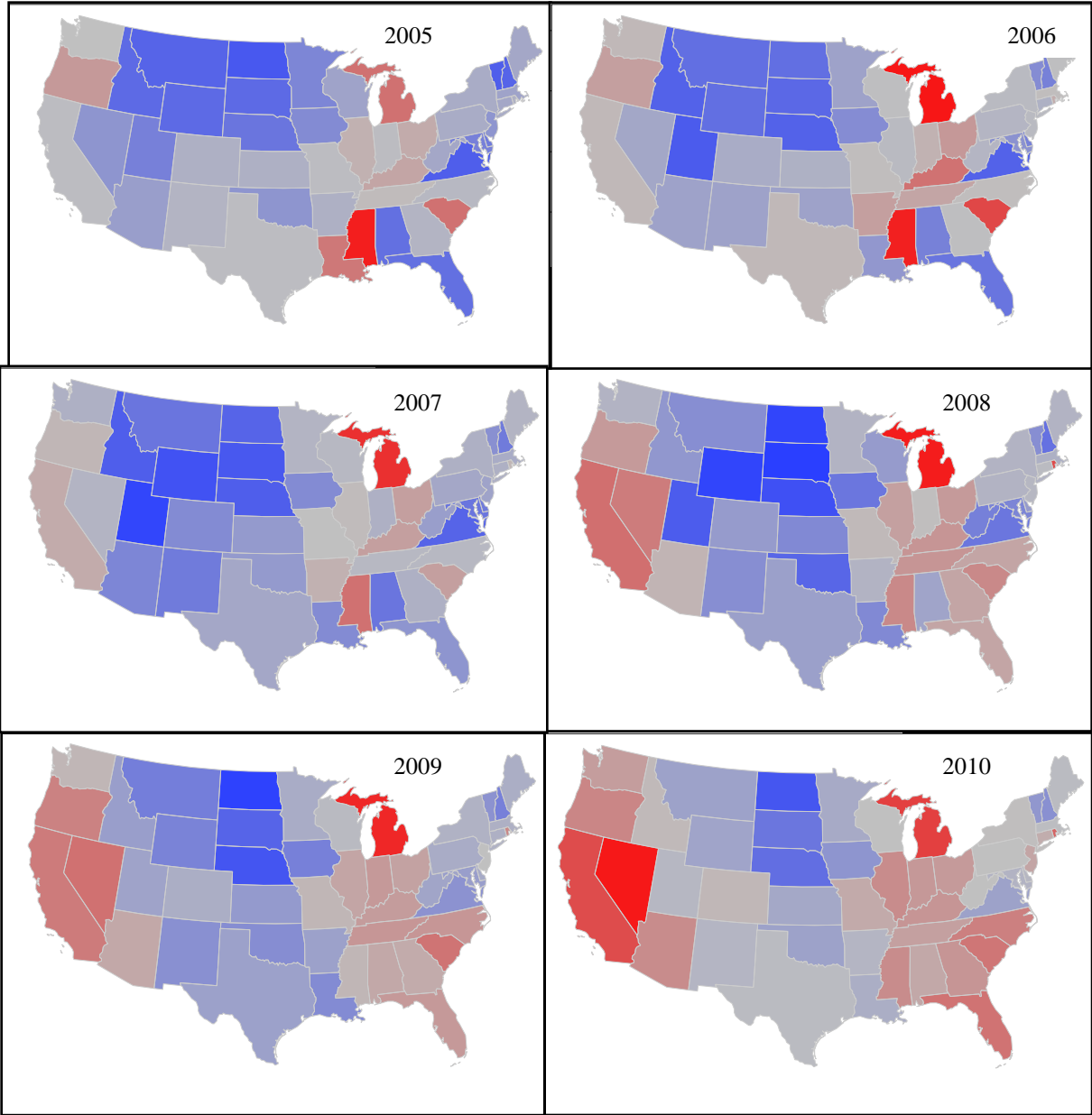
**Figure A.4.5: Flow Map over Years: Reefer Load non-Managed**



**Figure A.4.6: Flow Map - Total: Reefer Load non-Managed**

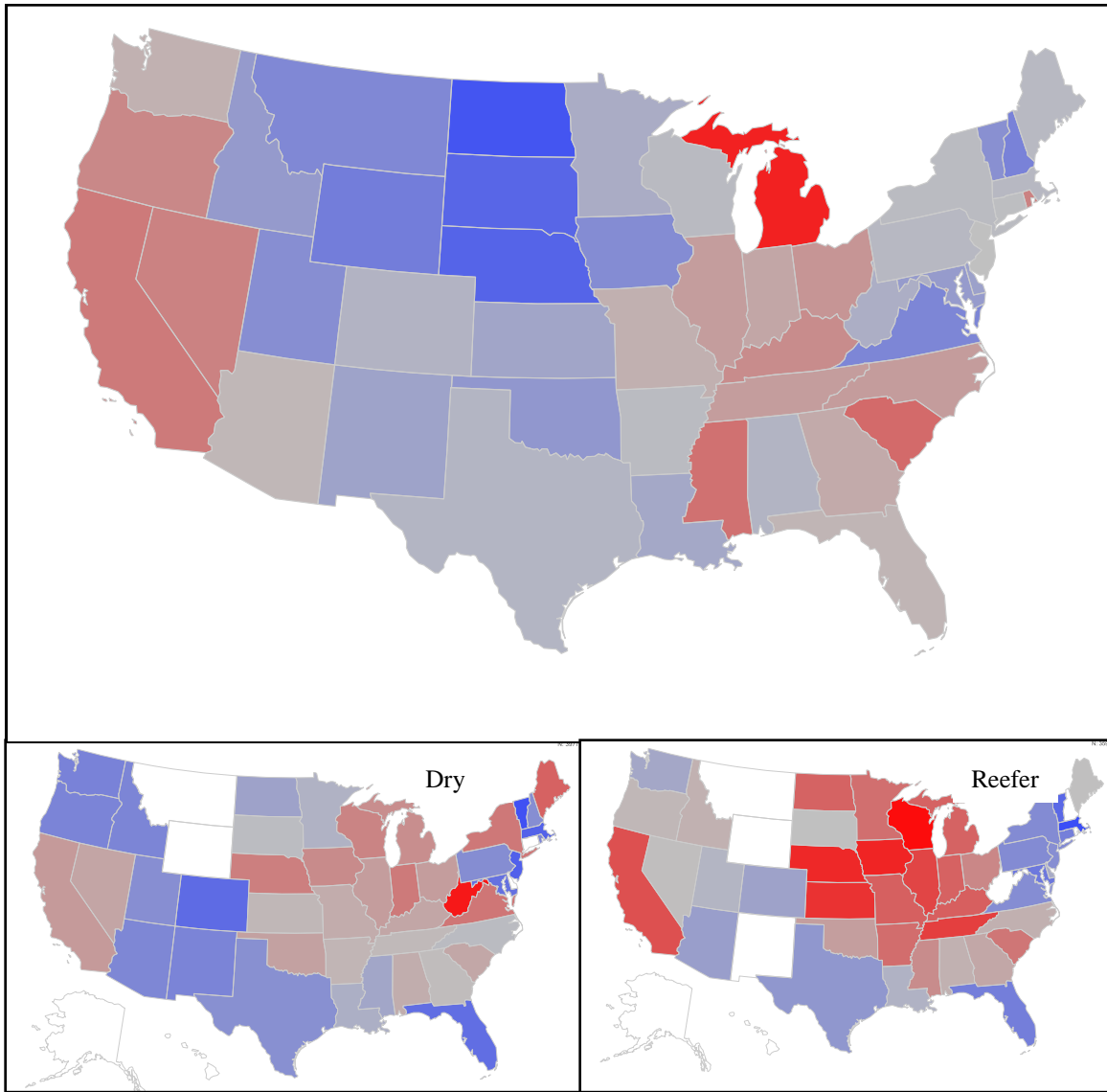


**A.5 U.S. Unemployment Maps**



**Figure A.5: U.S. Unemployment from 2005 to 2010**

## A.6 U.S. Unemployment vs. TL Prices



**Figure A.6: U.S. Unemployment v.s TL Prices**

Figure A.5.2 shows the average unemployment rate in the U.S. along with the exploratory maps for the dry and reefer product (managed lanes data). By visual inspection is noted that the TL prices follow the patterns for the dry commodities TL prices. The central north east of the U.S is dissimilar when compared with the reefer TL prices.



## APPENDIX B: Spatial Panel SAR and SDM

### B.1 SAR Fixed Effects

#### B.1.1 Dry Managed

R-squared      0.3034  
 corr-squared   0.1946  
 sigma<sup>2</sup>        0.0277

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.22047</b>	<b>-4.653999</b>	<b>0.000003</b>
<b>Unemployment</b>	<b>-0.254415</b>	<b>-3.938282</b>	<b>0.000082</b>
Fuel	-0.055605	-0.60771	0.54338
Population	-0.702584	-0.648634	0.516575
Gross State Product	0.436732	0.970359	0.331868
Wage	0.713826	1.344737	0.17871
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.789534</b>	<b>0.000002</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.2295</b>	<b>-4.8254</b>	<b>0.0496</b>	<b>3.3793</b>	<b>-0.1798</b>	<b>-4.7195</b>
<b>-0.2604</b>	<b>-3.9776</b>	<b>0.0568</b>	<b>2.863</b>	<b>-0.2037</b>	<b>-4.0381</b>
-0.054	-0.5778	0.0115	0.5555	-0.0425	-0.5785
-0.7479	-0.6912	0.1631	0.6696	-0.5848	-0.69
0.4549	0.9737	-0.0989	-0.9294	0.356	0.9734
0.724	1.3462	-0.1581	-1.256	0.5659	1.3478

#### B.1.2 Dry non-Managed

R-squared      0.2612  
 corr-squared   0.3241  
 sigma<sup>2</sup>        0.0323

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.20792</b>	<b>-3.704322</b>	<b>0.000212</b>
<b>Unemployment</b>	<b>-0.417512</b>	<b>-6.253914</b>	<b>0</b>
Fuel	-0.174104	-1.335529	0.181703
Population	1.754947	1.512088	0.130511
Gross State Product	0.077693	0.154908	0.876894
Wage	-0.323132	-0.572932	0.566691
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.896176</b>	<b>0.000001</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.2182</b>	<b>-3.7622</b>	<b>0.0478</b>	<b>3.0463</b>	<b>-0.1704</b>	<b>-3.6655</b>
<b>-0.4354</b>	<b>-6.1976</b>	<b>0.0963</b>	<b>3.4796</b>	<b>-0.3391</b>	<b>-6.4695</b>
-0.1883	-1.3947	0.0415	1.3119	-0.1469	-1.3907
1.8364	1.4951	-0.4031	-1.4159	1.4333	1.4854
0.0635	0.1281	-0.0137	-0.1255	0.0497	0.1281
-0.2949	-0.5316	0.0649	0.521	-0.23	-0.5305

### B.1.3 Reefer Managed

R-squared 0.3796  
 corr-squared 0.243  
 sigma^2 0.0148

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.152124</b>	<b>-5.03108</b>	<b>0</b>
Unemployment	-0.066626	-1.51579	0.129572
<b>Fuel</b>	<b>0.195023</b>	<b>3.082624</b>	<b>0.002052</b>
Population	-0.313099	-0.396457	0.691768
Gross State Product	-0.139073	-0.455153	0.648999
<b>Wage</b>	<b>0.995</b>	<b>2.857664</b>	<b>0.004268</b>
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.911202</b>	<b>0.000001</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.1568</b>	<b>-5.0795</b>	<b>0.034</b>	<b>3.4639</b>	<b>-0.1228</b>	<b>-5.076</b>
-0.0667	-1.4187	0.0144	1.3603	-0.0523	-1.4119
<b>0.203</b>	<b>3.1247</b>	<b>-0.0442</b>	<b>-2.516</b>	<b>0.1588</b>	<b>3.1431</b>
-0.324	-0.3985	0.0711	0.3961	-0.2529	-0.3963
-0.1271	-0.4116	0.0265	0.3949	-0.1006	-0.4133
<b>1.0032</b>	<b>2.7502</b>	<b>-0.2173</b>	<b>-2.3831</b>	<b>0.7859</b>	<b>2.736</b>

### B.1.4 Reefer non-Managed

R-squared 0.2349  
 corr-squared 0.2532  
 sigma^2 0.0178

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.205754</b>	<b>-4.594621</b>	<b>0.000004</b>
<b>Unemployment</b>	<b>-0.245861</b>	<b>-5.184134</b>	<b>0</b>
Fuel	0.058107	0.548953	0.583038
Population	-0.043397	-0.050433	0.959778
Gross State Product	0.291193	0.892973	0.371872
Wage	0.200135	0.538077	0.590524
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-5.006741</b>	<b>0.000001</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.2114</b>	<b>-4.4787</b>	<b>0.0452</b>	<b>3.223</b>	<b>-0.1662</b>	<b>-4.4714</b>
<b>-0.2536</b>	<b>-5.2045</b>	<b>0.0544</b>	<b>3.3938</b>	<b>-0.1992</b>	<b>-5.3054</b>
0.0605	0.5594	-0.013	-0.5377	0.0475	0.5613
-0.0111	-0.0123	0.0033	0.0171	-0.0078	-0.0109
0.2915	0.8584	-0.0632	-0.8374	0.2284	0.8563
0.1957	0.5039	-0.0416	-0.4859	0.1541	0.5054

## B.2 SAR Random Effects

### B.2.1 Dry Managed

R-squared 0.1497  
 corr-squared 0.2291  
 sigma<sup>2</sup> 0.0282

Variable	Coefficient	t-stat	z-probability
Intercept	-0.073447	-0.066581	0.946915
<b>Distance</b>	<b>-0.239623</b>	<b>-6.499916</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.21164</b>	<b>-4.877073</b>	<b>0.000001</b>
Fuel	0.008945	0.118218	0.905895
Population	0.140779	1.039124	0.298747
Gross State Product	-0.079902	-0.626081	0.531262
<b>Wage</b>	<b>0.47552</b>	<b>1.898635</b>	<b>0.057612</b>
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-5.019537</b>	<b>0.000001</b>
<b>teta</b>	<b>0.850427</b>	<b>7.199674</b>	<b>0</b>

direct	t-stat	indirect	t-stat	total	t-stat
-0.0261	-0.0218	0.0015	0.0056	-0.0246	-0.0262
<b>-0.2503</b>	<b>-6.466</b>	<b>0.0546</b>	<b>3.9336</b>	<b>-0.1957</b>	<b>-6.3385</b>
<b>-0.2186</b>	<b>-4.7972</b>	<b>0.0477</b>	<b>3.429</b>	<b>-0.1709</b>	<b>-4.7578</b>
0.0062	0.0822	-0.0017	-0.1009	0.0045	0.0762
0.1436	0.9528	-0.0309	-0.9341	0.1127	0.9474
-0.0807	-0.5644	0.0171	0.5517	-0.0636	-0.5635
<b>0.4839</b>	<b>1.8273</b>	<b>-0.1046</b>	<b>-1.7565</b>	<b>0.3793</b>	<b>1.8036</b>

### B.2.2 Dry non-Managed

R-squared 0.1112  
 corr-squared 0.2786  
 sigma<sup>2</sup> 0.0323

Variable	Coefficient	t-stat	z-probability
Intercept	0.126466	0.114257	0.909034
<b>Distance</b>	<b>-0.19706</b>	<b>-4.623877</b>	<b>0.000004</b>
<b>Unemployment</b>	<b>-0.335566</b>	<b>-7.680373</b>	<b>0</b>
Fuel	-0.033851	-0.324803	0.74533
<b>Population</b>	<b>0.31109</b>	<b>2.399361</b>	<b>0.016424</b>
<b>Gross State Product</b>	<b>-0.276053</b>	<b>-2.247798</b>	<b>0.024589</b>
Wage	0.332443	1.352561	0.176196
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.970733</b>	<b>0.000001</b>
<b>teta</b>	<b>0.996782</b>	<b>7.731628</b>	<b>0</b>

direct	t-stat	indirect	t-stat	total	t-stat
0.0874	0.0766	-0.0201	-0.078	0.0673	0.0757
<b>-0.2025</b>	<b>-4.5538</b>	<b>0.0446</b>	<b>3.3503</b>	<b>-0.1579</b>	<b>-4.4144</b>
<b>-0.352</b>	<b>-7.7272</b>	<b>0.0781</b>	<b>3.7977</b>	<b>-0.2739</b>	<b>-8.0548</b>
-0.0338	-0.3053	0.0072	0.2861	-0.0267	-0.3087
<b>0.3244</b>	<b>2.3759</b>	<b>-0.072</b>	<b>-2.0603</b>	<b>0.2524</b>	<b>2.3799</b>
<b>-0.2874</b>	<b>-2.2286</b>	<b>0.0637</b>	<b>1.9591</b>	<b>-0.2237</b>	<b>-2.2314</b>
0.3532	1.3976	-0.0774	-1.3344	0.2758	1.3886

### B.2.3 Reefer Managed

R-squared 0.2309  
 corr-squared 0.1781  
 sigma<sup>2</sup> 0.0153

Variable	Coefficient	t-stat	z-probability
Intercept	0.50951	0.550845	0.58174
<b>Distance</b>	<b>-0.180649</b>	<b>-6.550777</b>	<b>0</b>
Unemployment	-0.036465	-1.111966	0.266153
<b>Fuel</b>	<b>0.215913</b>	<b>3.862012</b>	<b>0.000112</b>
Population	0.105216	0.932509	0.351074
Gross State Product	-0.105314	-0.981392	0.3264
Wage	0.292152	1.44114	0.149545
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-5.016402</b>	<b>0.000001</b>
<b>teta</b>	<b>0.68939</b>	<b>7.072028</b>	<b>0</b>

direct	t-stat	indirect	t-stat	total	t-stat
0.5081	0.5346	-0.1096	-0.5325	0.3985	0.5313
<b>-0.1865</b>	<b>-6.6773</b>	<b>0.0399</b>	<b>3.847</b>	<b>-0.1465</b>	<b>-6.6709</b>
-0.0373	-1.0812	0.0079	1.0385	-0.0294	-1.0791
<b>0.2217</b>	<b>3.8572</b>	<b>-0.0477</b>	<b>-2.8504</b>	<b>0.174</b>	<b>3.9354</b>
0.1109	0.9736	-0.0237	-0.9517	0.0872	0.9689
-0.111	-1.0193	0.0237	0.9942	-0.0873	-1.0143
0.3051	1.4621	-0.065	-1.3928	0.2401	1.4558

### B.2.4 Reefer non-Managed

R-squared 0.062  
 corr-squared 0.2165  
 sigma<sup>2</sup> 0.0181

Variable	Asymptot	t-stat	z-probability
Intercept	1.241961	1.600861	0.109408
<b>Distance</b>	<b>-0.216468</b>	<b>-6.307851</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.182748</b>	<b>-5.802873</b>	<b>0</b>
Fuel	0.12556	1.42889	0.153036
Population	0.118127	1.309014	0.19053
Gross State Product	-0.091186	-1.070193	0.284533
Wage	0.117727	0.692013	0.488929
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-5.108726</b>	<b>0</b>
<b>teta</b>	<b>0.997291</b>	<b>8.050233</b>	<b>0</b>

direct	t-stat	indirect	t-stat	total	t-stat
1.3208	1.6946	-0.2861	-1.5776	1.0347	1.6935
<b>-0.2255</b>	<b>-6.2505</b>	<b>0.0488</b>	<b>3.8724</b>	<b>-0.1767</b>	<b>-6.229</b>
<b>-0.1886</b>	<b>-5.8234</b>	<b>0.041</b>	<b>3.6436</b>	<b>-0.1476</b>	<b>-5.9414</b>
0.1318	1.4195	-0.0285	-1.3573	0.1033	1.413
0.1211	1.3357	-0.0264	-1.2728	0.0947	1.3338
-0.0933	-1.0872	0.0203	1.0495	-0.073	-1.0849
0.1137	0.6491	-0.0245	-0.6308	0.0892	0.6492

### B.3 SDM Fixed Effects

#### B.3.1 Dry Managed

R-squared 0.3742  
 corr-squared 0.2952  
 sigma<sup>2</sup> 0.0249

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.20378</b>	<b>-4.39933</b>	<b>0.000011</b>
Unemployment	-0.098533	-0.82932	0.406921
Fuel	0.030361	0.236851	0.812773
Population	-0.953783	-0.81339	0.415996
Gross State Product	0.179779	0.304248	0.760939
Wage	0.447749	0.798865	0.424369
<b>W*Distance</b>	<b>-0.148356</b>	<b>-3.27506</b>	<b>0.001056</b>
W*Unemployment	-0.166974	-1.37952	0.167736
<b>W*Fuel</b>	<b>-0.21562</b>	<b>-1.66323</b>	<b>0.096267</b>
W*Population	0.942315	0.735144	0.462252
W*Gross State Product	0.910777	1.445034	0.148448
W*Wage	-0.737687	-1.3947	0.163106
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.63881</b>	<b>0.000004</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.2103</b>	<b>-4.2284</b>	<b>0.0459</b>	<b>2.881</b>	<b>-0.1644</b>	<b>-4.3292</b>
-0.1012	-0.8297	0.0223	0.7987	-0.0789	-0.8282
0.0302	0.2315	-0.0064	-0.2205	0.0238	0.2328
-0.9529	-0.779	0.204	0.7365	-0.7488	-0.7823
0.1497	0.2389	-0.034	-0.2464	0.1157	0.235
0.4345	0.7599	-0.0926	-0.7248	0.3418	0.7614
<b>-0.1538</b>	<b>-3.3438</b>	<b>0.0337</b>	<b>2.5284</b>	<b>-0.1201</b>	<b>-3.4108</b>
-0.1752	-1.442	0.0378	1.3453	-0.1373	-1.4403
<b>-0.2195</b>	<b>-1.7058</b>	<b>0.0474</b>	<b>1.5792</b>	<b>-0.1721</b>	<b>-1.6956</b>
0.9937	0.7374	-0.2135	-0.6988	0.7802	0.7412
0.9515	1.424	-0.2057	-1.3401	0.7458	1.4178
-0.7287	-1.4224	0.1576	1.3373	-0.5711	-1.4169

#### B.3.2 Dry non-Managed

R-squared 0.3169  
 corr-squared 0.3988  
 sigma<sup>2</sup> 0.0298

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.219151</b>	<b>-3.93598</b>	<b>0.000083</b>
<b>Unemployment</b>	<b>-0.249017</b>	<b>-1.92824</b>	<b>0.053825</b>
Fuel	-0.069996	-0.46032	0.645286
Population	0.79159	0.59775	0.550007
Gross State Product	0.263529	0.410662	0.68132
Wage	-0.006387	-0.01044	0.991672
W*Distance	-0.079206	-1.49901	0.133872
W*Unemployment	-0.181655	-1.3541	0.175706
W*Fuel	-0.187176	-1.27736	0.201474
W*Population	2.263892	1.668156	0.095285
W*Gross State Product	0.304838	0.449818	0.652842
<b>W*Wage</b>	<b>-1.294176</b>	<b>-2.34752</b>	<b>0.018899</b>
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.73299</b>	<b>0.000002</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.2308</b>	<b>-4.0188</b>	<b>0.051</b>	<b>2.8949</b>	<b>-0.1799</b>	<b>-4.0219</b>
<b>-0.2524</b>	<b>-1.8879</b>	<b>0.0558</b>	<b>1.7014</b>	<b>-0.1967</b>	<b>-1.8862</b>
-0.0703	-0.4513	0.0155	0.445	-0.0549	-0.4493
0.8044	0.5886	-0.1784	-0.575	0.626	0.5868
0.2996	0.4661	-0.0664	-0.4591	0.2332	0.4629
0.0297	0.046	-0.0036	-0.0252	0.0261	0.0516
-0.0782	-1.3958	0.0177	1.286	-0.0605	-1.4014
-0.1991	-1.4718	0.0444	1.3482	-0.1547	-1.4761
-0.1985	-1.3232	0.0443	1.2426	-0.1541	-1.3212
2.363	1.6841	-0.5272	-1.4936	1.8358	1.6984
0.2718	0.389	-0.0613	-0.3844	0.2105	0.3867
<b>-1.3335</b>	<b>-2.2704</b>	<b>0.2939</b>	<b>2.005</b>	<b>-1.0395</b>	<b>-2.2556</b>

### B.3.3 Reefer Managed

R-squared 0.4088  
 corr-squared 0.282  
 sigma<sup>2</sup> 0.0141

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.126316</b>	<b>-3.83619</b>	<b>0.000125</b>
Unemployment	-0.077309	-0.86635	0.386297
<b>Fuel</b>	<b>0.354889</b>	<b>3.009801</b>	<b>0.002614</b>
Population	-0.068395	-0.07156	0.942949
Gross State Product	-0.369161	-0.95665	0.338744
<b>Wage</b>	<b>0.761716</b>	<b>2.084529</b>	<b>0.037112</b>
W*Distance	-0.038224	-1.15149	0.249529
W*Unemployment	-0.03095	-0.33561	0.737168
W*Fuel	-0.212497	-1.81311	0.069815
W*Population	-0.487759	-0.51144	0.609043
W*Gross State Product	0.120759	0.276658	0.782043
<b>W*Wage</b>	<b>0.732928</b>	<b>1.938061</b>	<b>0.052616</b>
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.74422</b>	<b>0.000002</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.1307</b>	<b>-3.7084</b>	<b>0.0283</b>	<b>2.7574</b>	<b>-0.1024</b>	<b>-3.7562</b>
-0.0796	-0.8517	0.0172	0.8154	-0.0624	-0.8527
<b>0.372</b>	<b>3.0044</b>	<b>-0.0797</b>	<b>-2.56</b>	<b>0.2923</b>	<b>2.9584</b>
-0.0845	-0.0856	0.0191	0.0889	-0.0654	-0.0841
-0.367	-0.9194	0.0784	0.8889	-0.2886	-0.9173
<b>0.8096</b>	<b>2.1147</b>	<b>-0.1753</b>	<b>-1.877</b>	<b>0.6343</b>	<b>2.1135</b>
-0.0378	-1.0898	0.0083	1.032	-0.0295	-1.0916
-0.0324	-0.3416	0.0071	0.3394	-0.0253	-0.3397
-0.2232	-1.8676	0.0474	1.7387	-0.1757	-1.8506
-0.4996	-0.5142	0.107	0.4995	-0.3926	-0.5142
0.0973	0.213	-0.0195	-0.1964	0.0777	0.216
<b>0.7538</b>	<b>1.9215</b>	<b>-0.1633</b>	<b>-1.7095</b>	<b>0.5905</b>	<b>1.9319</b>

### B.3.4 Reefer non-Managed

R-squared 0.2753  
 corr-squared 0.3074  
 sigma<sup>2</sup> 0.0168

Variable	Coefficient	t-stat	z-probability
<b>Distance</b>	<b>-0.187728</b>	<b>-4.04876</b>	<b>0.000051</b>
Unemployment	-0.084783	-0.89009	0.373419
Fuel	0.003661	0.033765	0.973065
Population	-0.222317	-0.21921	0.826489
Gross State Product	0.168923	0.416907	0.676746
Wage	0.104307	0.268714	0.788149
W*Distance	-0.046256	-1.11837	0.263408
<b>W*Unemployment</b>	<b>-0.170526</b>	<b>-1.75407</b>	<b>0.079419</b>
W*Fuel	0.132494	1.256491	0.208938
W*Population	0.341755	0.340987	0.733113
W*Gross State Product	0.247195	0.537902	0.590645
W*Wage	-0.149145	-0.36574	0.714562
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.77879</b>	<b>0.000002</b>

direct	t-stat	indirect	t-stat	total	t-stat
<b>-0.1918</b>	<b>-3.9498</b>	<b>0.0412</b>	<b>2.9837</b>	<b>-0.1506</b>	<b>-3.9011</b>
-0.0846	-0.8697	0.0183	0.8449	-0.0663	-0.8655
0.0034	0.0308	-0.0007	-0.0308	0.0026	0.0307
-0.2207	-0.2142	0.0495	0.2157	-0.1712	-0.2124
0.178	0.424	-0.0391	-0.4232	0.1389	0.4208
0.1035	0.2606	-0.0223	-0.2565	0.0812	0.2599
-0.0469	-1.0768	0.0105	1.0241	-0.0364	-1.078
<b>-0.1766</b>	<b>-1.7558</b>	<b>0.0384</b>	<b>1.6166</b>	<b>-0.1382</b>	<b>-1.7556</b>
0.1383	1.3149	-0.0296	-1.2429	0.1087	1.309
0.3399	0.3323	-0.0743	-0.3341	0.2655	0.3296
0.2438	0.5241	-0.0523	-0.5135	0.1916	0.5228
-0.1462	-0.3393	0.0309	0.3209	-0.1152	-0.342

## B.4 SDM Random Effects

### B.4.1 Dry Managed

R-squared 0.2016; corr-squared 0.2987; sigma<sup>2</sup> 0.0265

Variable	Coefficient	t-stat	z-probability	direct	t-stat	indirect	t-stat	total	t-stat
Intercept	-0.736462	-0.46078	0.644954	-0.8243	-0.4891	0.1836	0.4797	-0.6407	-0.4877
<b>Distance</b>	<b>-0.19821</b>	<b>-5.34647</b>	<b>0</b>	<b>-0.2052</b>	<b>-5.2577</b>	<b>0.0452</b>	<b>3.3428</b>	<b>-0.16</b>	<b>-5.2558</b>
<b>Unemployment</b>	<b>-0.202605</b>	<b>-2.79276</b>	<b>0.005226</b>	<b>-0.2098</b>	<b>-2.7281</b>	<b>0.0461</b>	<b>2.3073</b>	<b>-0.1636</b>	<b>-2.709</b>
Fuel	0.075013	0.610383	0.541608	0.0853	0.6554	-0.0185	-0.638	0.0668	0.654
Population	0.089933	0.615453	0.538255	0.0945	0.6125	-0.0212	-0.5972	0.0734	0.6107
Gross State Product	-0.04164	-0.30438	0.760837	-0.0443	-0.3043	0.0101	0.3043	-0.0342	-0.3019
<b>Wage</b>	<b>0.502256</b>	<b>2.119904</b>	<b>0.034014</b>	<b>0.5196</b>	<b>2.0913</b>	<b>-0.1149</b>	<b>-1.8435</b>	<b>0.4048</b>	<b>2.0898</b>
<b>W*Distance</b>	<b>-0.123769</b>	<b>-3.28846</b>	<b>0.001007</b>	<b>-0.1288</b>	<b>-3.2636</b>	<b>0.0288</b>	<b>2.3707</b>	<b>-0.1</b>	<b>-3.4054</b>
W*Unemployment	-0.042499	-0.54521	0.585609	-0.045	-0.5478	0.0104	0.5591	-0.0346	-0.5399
W*Fuel	-0.114234	-0.92986	0.352443	-0.1257	-0.9749	0.0275	0.9259	-0.0982	-0.9768
W*Population	0.154454	1.068044	0.2855	0.164	1.076	-0.0362	-1.0346	0.1278	1.0723
W*Gross State Product	-0.1124	-0.83001	0.406531	-0.1194	-0.8412	0.0262	0.8175	-0.0931	-0.8385
W*Wage	0.191813	0.7287	0.466185	0.2118	0.7801	-0.0475	-0.7627	0.1643	0.7774
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.60581</b>	<b>0.000004</b>						
<b>teta</b>	<b>0.890872</b>	<b>7.31208</b>	<b>0</b>						

### B.4.2 Dry non-Managed

R-squared 0.1488; corr-squared 0.3211; sigma<sup>2</sup> 0.031

Variable	Coefficient	t-stat	z-probability	direct	t-stat	indirect	t-stat	total	t-stat
Intercept	0.744967	0.448732	0.653625	0.7649	0.4561	-0.1704	-0.4494	0.5944	0.4542
<b>Distance</b>	<b>-0.179659</b>	<b>-4.11344</b>	<b>0.000039</b>	<b>-0.1866</b>	<b>-3.9422</b>	<b>0.0414</b>	<b>2.9699</b>	<b>-0.1452</b>	<b>-3.9024</b>
<b>Unemployment</b>	<b>-0.242802</b>	<b>-3.20033</b>	<b>0.001373</b>	<b>-0.2516</b>	<b>-2.9889</b>	<b>0.0559</b>	<b>2.4513</b>	<b>-0.1956</b>	<b>-2.9811</b>
Fuel	0.01821	0.133754	0.893597	0.0243	0.1716	-0.006	-0.1876	0.0182	0.1657
Population	0.205003	1.470159	0.141519	0.211	1.4835	-0.0467	-1.3837	0.1643	1.483
Gross State Product	-0.181214	-1.38014	0.167544	-0.1868	-1.3879	0.0413	1.3048	-0.1455	-1.3863
Wage	0.335091	1.357177	0.174725	0.3394	1.3044	-0.0748	-1.2489	0.2646	1.2964
W*Distance	-0.067849	-1.50451	0.13245	-0.0722	-1.5429	0.0164	1.4009	-0.0559	-1.5542
W*Unemployment	-0.101553	-1.22103	0.222075	-0.1083	-1.1765	0.0245	1.1123	-0.0838	-1.1776
W*Fuel	-0.04049	-0.31314	0.754175	-0.0467	-0.3478	0.0108	0.3522	-0.0359	-0.3439
W*Population	0.078869	0.52062	0.602632	0.0892	0.559	-0.0205	-0.5597	0.0687	0.5537
W*Gross State Product	-0.037267	-0.26111	0.794009	-0.0451	-0.3012	0.0107	0.3147	-0.0344	-0.295
W*Wage	-0.194701	-0.72203	0.470275	-0.192	-0.6783	0.0418	0.6531	-0.1501	-0.6789
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.71998</b>	<b>0.000002</b>						
<b>teta</b>	<b>0.997133</b>	<b>7.732233</b>	<b>0</b>						

### B.4.3 Reefer Managed

R-squared 0.264; corr-squared 0.2549; sigma<sup>2</sup> 0.0146

Variable	Coefficient	t-stat	z-probability	direct	t-stat	indirect	t-stat	total	t-stat
Intercept	-0.827157	-0.62161	0.534196	-0.8012	-0.5804	0.1679	0.5575	-0.6333	-0.5823
<b>Distance</b>	<b>-0.152088</b>	<b>-5.20899</b>	<b>0</b>	<b>-0.1578</b>	<b>-5.2519</b>	<b>0.034</b>	<b>3.4685</b>	<b>-0.1238</b>	<b>-5.269</b>
Unemployment	-0.106485	-1.76476	0.077605	-0.1087	-1.7494	0.0233	1.6168	-0.0854	-1.7468
<b>Fuel</b>	<b>0.296684</b>	<b>2.660288</b>	<b>0.007807</b>	<b>0.3111</b>	<b>2.817</b>	<b>-0.0667</b>	<b>-2.4246</b>	<b>0.2444</b>	<b>2.799</b>
Population	0.205493	1.680035	0.09295	0.2099	1.6614	-0.0451	-1.5279	0.1649	1.6646
Gross State Product	-0.201986	-1.75706	0.078908	-0.206	-1.7342	0.0442	1.5886	-0.1618	-1.7368
Wage	0.254233	1.296983	0.194637	0.251	1.2582	-0.054	-1.1976	0.197	1.256
<b>W*Distance</b>	<b>-0.062007</b>	<b>-2.08707</b>	<b>0.036881</b>	<b>-0.0626</b>	<b>-1.9935</b>	<b>0.0138</b>	<b>1.7324</b>	<b>-0.0488</b>	<b>-2.0278</b>
W*Unemployment	0.030427	0.465944	0.641255	0.0319	0.4648	-0.0068	-0.4557	0.0251	0.4639
W*Fuel	-0.124741	-1.12117	0.262216	-0.134	-1.2118	0.0283	1.1573	-0.1057	-1.2087
W*Population	0.044247	0.33773	0.735567	0.0416	0.3053	-0.0089	-0.2957	0.0327	0.3062
W*Gross State Product	-0.041587	-0.34191	0.732418	-0.039	-0.3056	0.0084	0.2969	-0.0306	-0.3062
<b>W*Wage</b>	<b>0.4337</b>	<b>1.952573</b>	<b>0.05087</b>	<b>0.444</b>	<b>1.9401</b>	<b>-0.0949</b>	<b>-1.7994</b>	<b>0.3491</b>	<b>1.9264</b>
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.74401</b>	<b>0.000002</b>						
<b>teta</b>	<b>0.724526</b>	<b>7.160024</b>	<b>0</b>						

### B.4.4 Reefer non-Managed

R-squared 0.1432; corr-squared 0.3029; sigma<sup>2</sup> 0.0166

Variable	Coefficient	t-stat	z-probability	direct	t-stat	indirect	t-stat	total	t-stat
Intercept	0.586764	0.516259	0.605674	0.5919	0.4877	-0.1303	-0.4864	0.4617	0.4843
<b>Distance</b>	<b>-0.180714</b>	<b>-5.01289</b>	<b>0.000001</b>	<b>-0.1883</b>	<b>-5.0415</b>	<b>0.0407</b>	<b>3.4324</b>	<b>-0.1475</b>	<b>-4.9922</b>
<b>Unemployment</b>	<b>-0.125029</b>	<b>-2.35029</b>	<b>0.018759</b>	<b>-0.1286</b>	<b>-2.4089</b>	<b>0.0279</b>	<b>2.1054</b>	<b>-0.1007</b>	<b>-2.4071</b>
Fuel	0.097077	1.059184	0.289516	0.0965	1.0397	-0.0211	-1.016	0.0753	1.0338
Population	0.132367	1.330473	0.183362	0.1384	1.3259	-0.0299	-1.2589	0.1085	1.3236
Gross State Product	-0.114709	-1.23218	0.21788	-0.1199	-1.2331	0.0259	1.1777	-0.094	-1.2306
Wage	0.077271	0.46642	0.640915	0.0849	0.4884	-0.0187	-0.485	0.0663	0.4856
<b>W*Distance</b>	<b>-0.068316</b>	<b>-2.00156</b>	<b>0.045332</b>	<b>-0.0695</b>	<b>-1.9404</b>	<b>0.0154</b>	<b>1.6923</b>	<b>-0.0541</b>	<b>-1.9731</b>
W*Unemployment	-0.086191	-1.50107	0.133338	-0.0907	-1.6104	0.0198	1.4949	-0.0709	-1.6095
<b>W*Fuel</b>	<b>0.163515</b>	<b>1.908255</b>	<b>0.056358</b>	<b>0.1724</b>	<b>1.8883</b>	<b>-0.0373</b>	<b>-1.7192</b>	<b>0.1352</b>	<b>1.8875</b>
W*Population	0.085865	0.798923	0.424335	0.0868	0.7819	-0.0188	-0.7609	0.068	0.7793
W*Gross State Product	-0.047916	-0.47924	0.631772	-0.0487	-0.4716	0.0106	0.4646	-0.0381	-0.4697
W*Wage	0.088792	0.461279	0.644598	0.098	0.4786	-0.0214	-0.472	0.0766	0.4767
<b>W*dep.var.</b>	<b>-0.236068</b>	<b>-4.79035</b>	<b>0.000002</b>						
<b>teta</b>	<b>0.997204</b>	<b>8.049504</b>	<b>0</b>						



## APPENDIX C: Spatial Panel SEM and SDEM

### C.1 SEM Fixed Effects

#### C.3.1 Dry Managed

R-squared 0.4259  
 corr-squared 0.2215  
 sigma<sup>2</sup> 0.0201  
 log-likelihood 126.23746

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.118874</b>	<b>-2.925687</b>	<b>0.003437</b>
<b>Unemployment</b>	<b>-0.189078</b>	<b>-3.063183</b>	<b>0.00219</b>
Fuel	0.041172	0.456251	0.64821
Population	-0.913571	-0.911337	0.362118
Gross State Product	0.008803	0.021589	0.982776
<b>Wage</b>	<b>1.111834</b>	<b>2.428572</b>	<b>0.015158</b>
<b>spat.aut.</b>	<b>0.233979</b>	<b>4.538282</b>	<b>0.000006</b>

#### C.3.2 Dry non-Managed

R-squared 0.4191  
 corr-squared 0.3438  
 sigma<sup>2</sup> 0.02  
 log-likelihood 127.29688

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.155377</b>	<b>-3.670674</b>	<b>0.000242</b>
<b>Unemployment</b>	<b>-0.322753</b>	<b>-5.346767</b>	<b>0</b>
Fuel	-0.078129	-0.666957	0.504799
Population	0.637488	0.619715	0.535445
Gross State Product	-0.280867	-0.658947	0.50993
Wage	0.581294	1.305896	0.191588
<b>spat.aut.</b>	<b>0.312982</b>	<b>6.580618</b>	<b>0</b>

### C.3.3 Reefer Managed

R-squared 0.4798  
 corr-squared 0.2541  
 sigma<sup>2</sup> 0.0118  
 log-likelihood 200.4651

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.135249</b>	<b>-4.83798</b>	<b>0.000001</b>
Unemployment	-0.029351	-0.689588	0.490453
<b>Fuel</b>	<b>0.177397</b>	<b>2.828569</b>	<b>0.004676</b>
Population	-0.330062	-0.43685	0.66222
Gross State Product	-0.11946	-0.427017	0.669367
<b>Wage</b>	<b>0.628239</b>	<b>2.005694</b>	<b>0.044889</b>
<b>spat.aut.</b>	<b>0.150997</b>	<b>2.895615</b>	<b>0.003784</b>

### C.3.4 Reefer non-Managed

R-squared 0.3893  
 corr-squared 0.2729  
 sigma<sup>2</sup> 0.0117  
 log-likelihood 202.02521

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.179204</b>	<b>-4.74031</b>	<b>0.000002</b>
<b>Unemployment</b>	<b>-0.186152</b>	<b>-4.137153</b>	<b>0.000035</b>
Fuel	-0.064033	-0.723303	0.469494
Population	0.383944	0.486886	0.626339
Gross State Product	0.049971	0.181331	0.856108
Wage	0.235128	0.776312	0.437565
<b>spat.aut.</b>	<b>0.298997</b>	<b>6.294831</b>	<b>0</b>

## C.2 SEM Random Effects

### C.2.1 Dry Managed

R-squared 0.3687  
 corr-squared 0.262  
 sigma<sup>2</sup> 0.0209  
 log-likelihood 101.95452

Variable	Asymptot	t-stat	z-probability
Intercept	0.9535	0.922593	0.356219
<b>Distance</b>	<b>-0.18319</b>	<b>-5.512344</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.149604</b>	<b>-3.575248</b>	<b>0.00035</b>
Fuel	0.029953	0.400654	0.688675
Population	-0.019704	-0.143894	0.885584
Gross State Product	0.051258	0.390176	0.696406
Wage	0.213382	1.015601	0.309819
<b>spat.aut.</b>	<b>0.186906</b>	<b>3.43558</b>	<b>0.000591</b>
teta	0.041319	1.018699	0.308346

### C.2.2 Dry non-Managed

R-squared 0.4627  
 corr-squared 0.2997  
 sigma<sup>2</sup> 0.0195  
 log-likelihood 108.39845

Variable	Asymptot	t-stat	z-probability
Intercept	-0.947713	-0.956645	0.338746
<b>Distance</b>	<b>-0.167898</b>	<b>-4.929651</b>	<b>0.000001</b>
<b>Unemployment</b>	<b>-0.248989</b>	<b>-5.933773</b>	<b>0</b>
Fuel	-0.032417	-0.329086	0.742091
<b>Population</b>	<b>0.353201</b>	<b>2.832133</b>	<b>0.004624</b>
<b>Gross State Product</b>	<b>-0.323811</b>	<b>-2.703456</b>	<b>0.006862</b>
<b>Wage</b>	<b>0.521593</b>	<b>2.50859</b>	<b>0.012121</b>
<b>spat.aut.</b>	<b>0.301832</b>	<b>6.277132</b>	<b>0</b>
teta	0.037572	0.983639	0.325293

### C.2.3 Reefer Managed

R-squared 0.3968  
 corr-squared 0.1996  
 sigma^2 0.012  
 log-likelihood 160.59395

Variable	Asymptot	t-stat	z-probability
<b>Intercept</b>	<b>1.805385</b>	<b>2.008202</b>	<b>0.044622</b>
<b>Distance</b>	<b>-0.148121</b>	<b>-5.600937</b>	<b>0</b>
Unemployment	0.001727	0.051385	0.959018
<b>Fuel</b>	<b>0.195909</b>	<b>3.453846</b>	<b>0.000553</b>
Population	-0.012917	-0.110967	0.911642
Gross State Product	-0.014966	-0.130862	0.895884
Wage	-0.040163	-0.20352	0.838728
<b>spat.aut.</b>	<b>0.173692</b>	<b>3.123657</b>	<b>0.001786</b>
<b>teta</b>	<b>0.269253</b>	<b>3.432726</b>	<b>0.000598</b>

### C.2.4 Reefer non-Managed

R-squared 0.3892  
 corr-squared 0.2322  
 sigma^2 0.0118  
 log-likelihood 171.20695

Variable	Asymptot	t-stat	z-probability
<b>Intercept</b>	<b>2.275139</b>	<b>2.895183</b>	<b>0.003789</b>
<b>Distance</b>	<b>-0.169272</b>	<b>-5.180506</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.102317</b>	<b>-3.004022</b>	<b>0.002664</b>
Fuel	-0.019301	-0.240093	0.810258
Population	-0.070522	-0.725241	0.468304
Gross State Product	0.067077	0.712138	0.476379
Wage	-0.057226	-0.338116	0.735276
<b>spat.aut.</b>	<b>0.311743</b>	<b>6.467401</b>	<b>0</b>
<b>teta</b>	<b>0.108655</b>	<b>2.224955</b>	<b>0.026084</b>

### C.3 SDEM Fixed Effects

#### C.3.1 Dry Managed

R-squared 0.4891

corr-squared 0.3053

sigma<sup>2</sup> 0.0183; log-likelihood 136.59518

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.154645</b>	<b>-3.82752</b>	<b>0.000129</b>
Unemployment	-0.034705	-0.380357	0.70368
Fuel	0.034296	0.344608	0.730389
Population	-1.015141	-1.043856	0.296552
Gross State Product	0.101118	0.215367	0.829481
Wage	0.683958	1.41156	0.15808
<b>W*Distance</b>	<b>-0.102785</b>	<b>-2.639899</b>	<b>0.008293</b>
<b>W*Unemployment</b>	<b>-0.186288</b>	<b>-1.937623</b>	<b>0.052669</b>
W*Fuel	-0.141337	-1.450028	0.147051
W*Population	1.05044	0.972909	0.330598
W*Gross State Product	0.638316	1.287381	0.197961
W*Wage	-0.702598	-1.532958	0.125286
<b>spat.aut.</b>	<b>0.214998</b>	<b>4.126812</b>	<b>0.000037</b>

#### C.3.2 Dry non-Managed

R-squared 0.4716

corr-squared 0.399

sigma<sup>2</sup> 0.0182; log-likelihood 137.42636

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.207945</b>	<b>-4.434371</b>	<b>0.000009</b>
Unemployment	-0.14891	-1.666582	0.095598
Fuel	-0.043496	-0.379626	0.704223
Population	0.722369	0.715089	0.474554
Gross State Product	0.328231	0.697351	0.485583
Wage	0.266477	0.538968	0.589909
W*Distance	-0.041332	-0.943393	0.34548
W*Unemployment	-0.160104	-1.640843	0.10083
W*Fuel	-0.12402	-1.111385	0.266403
W*Population	1.38615	1.317186	0.187776
W*Gross State Product	0.120692	0.246945	0.804951
<b>W*Wage</b>	<b>-1.337457</b>	<b>-2.98779</b>	<b>0.00281</b>
<b>spat.aut.</b>	<b>0.307968</b>	<b>6.449451</b>	<b>0</b>

### C.3.3 Reefer Managed

R-squared 0.5024

corr-squared 0.2865

sigma<sup>2</sup> 0.0113; log-likelihood 205.45475

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.131541</b>	<b>-4.565083</b>	<b>0.000005</b>
Unemployment	-0.052489	-0.699174	0.484443
<b>Fuel</b>	<b>0.36317</b>	<b>3.74445</b>	<b>0.000181</b>
Population	-0.362249	-0.438056	0.661345
Gross State Product	-0.155097	-0.469425	0.638766
<b>Wage</b>	<b>0.610621</b>	<b>1.874341</b>	<b>0.060883</b>
W*Distance	-0.008133	-0.283869	0.776511
W*Unemployment	-0.001224	-0.015482	0.987647
<b>W*Fuel</b>	<b>-0.25254</b>	<b>-2.62439</b>	<b>0.00868</b>
W*Population	-0.343963	-0.413464	0.679267
W*Gross State Product	0.053997	0.144223	0.885324
W*Wage	0.369345	1.094203	0.273866
<b>spat.aut.</b>	<b>0.151975</b>	<b>2.915416</b>	<b>0.003552</b>

### C.3.4 Reefer non-Managed

R-squared 0.4253

corr-squared 0.3117

sigma<sup>2</sup> 0.011; log-likelihood 209.63297

Variable	Asymptot	t-stat	z-probability
<b>Distance</b>	<b>-0.176559</b>	<b>-4.595427</b>	<b>0.000004</b>
Unemployment	-0.067173	-0.9773	0.328421
Fuel	-0.042454	-0.471643	0.637181
Population	0.304743	0.377623	0.70571
Gross State Product	0.013137	0.042046	0.966462
Wage	0.142118	0.444824	0.656447
W*Distance	0.015197	0.445301	0.656102
W*Unemployment	-0.123511	-1.67339	0.094251
<b>W*Fuel</b>	<b>0.172732</b>	<b>2.007867</b>	<b>0.044657</b>
W*Population	-0.528151	-0.646827	0.517744
W*Gross State Product	0.2243	0.640345	0.521948
W*Wage	0.004713	0.014166	0.988698
<b>spat.aut.</b>	<b>0.300997</b>	<b>6.34692</b>	<b>0</b>

## C.4 SDEM Random Effects

### C.4.1 Dry Managed

R-squared 0.3962; corr-squared 0.2961; sigma<sup>2</sup> 0.02; log-likelihood 107.78658

Variable	Asymptot	t-stat	z-probability
Intercept	0.322156	0.201415	0.840374
<b>Distance</b>	<b>-0.177843</b>	<b>-5.389978</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.173169</b>	<b>-3.076798</b>	<b>0.002092</b>
Fuel	0.08181	0.83218	0.405308
Population	-0.099478	-0.661508	0.508287
Gross State Product	0.125768	0.873706	0.382278
Wage	0.320021	1.516189	0.129471
<b>W*Distance</b>	<b>-0.088176</b>	<b>-2.792843</b>	<b>0.005225</b>
W*Unemployment	-0.003313	-0.053268	0.957518
W*Fuel	-0.099736	-1.040116	0.298286
W*Population	0.176513	1.149521	0.250341
W*Gross State Product	-0.152665	-1.058132	0.289995
W*Wage	0.149166	0.613857	0.53931
<b>spat.aut.</b>	<b>0.172557</b>	<b>3.161612</b>	<b>0.001569</b>
teta	0.031764	0.819686	0.412395

### C.4.2 Dry non-Managed

R-squared 0.4799; corr-squared 0.3188; sigma<sup>2</sup> 0.0189; log-likelihood 111.69086

Variable	Asymptot	t-stat	z-probability
Intercept	-0.985962	-0.625722	0.531497
<b>Distance</b>	<b>-0.174359</b>	<b>-4.800857</b>	<b>0.000002</b>
<b>Unemployment</b>	<b>-0.186306</b>	<b>-3.486741</b>	<b>0.000489</b>
Fuel	-0.024039	-0.237624	0.812173
<b>Population</b>	<b>0.303216</b>	<b>2.337349</b>	<b>0.019421</b>
<b>Gross State Product</b>	<b>-0.275856</b>	<b>-2.215333</b>	<b>0.026737</b>
<b>Wage</b>	<b>0.537648</b>	<b>2.487098</b>	<b>0.012879</b>
W*Distance	-0.022879	-0.614422	0.538936
W*Unemployment	-0.09993	-1.601857	0.109187
W*Fuel	-0.010642	-0.11255	0.910387
W*Population	0.103698	0.759776	0.447389
W*Gross State Product	-0.07874	-0.609485	0.542203
W*Wage	-0.067588	-0.284147	0.776298
<b>spat.aut.</b>	<b>0.304234</b>	<b>6.339129</b>	<b>0</b>
teta	0.038827	1.010093	0.312451

### C.4.3 Reefer Managed

R-squared 0.4136; corr-squared 0.249; sigma<sup>2</sup>

0.0117; log-likelihood 167.37376

Variable	Asymptot	t-stat	z-probability
Intercept	0.688963	0.51591	0.605918
<b>Distance</b>	<b>-0.141217</b>	<b>-5.317198</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.110123</b>	<b>-2.14222</b>	<b>0.032176</b>
<b>Fuel</b>	<b>0.324396</b>	<b>3.520932</b>	<b>0.00043</b>
Population	0.146884	1.080676	0.279841
Gross State Product	-0.157956	-1.206966	0.227445
Wage	-0.043421	-0.231431	0.81698
W*Distance	-0.021703	-0.856252	0.391858
<b>W*Unemployment</b>	<b>0.099281</b>	<b>1.799325</b>	<b>0.071967</b>
<b>W*Fuel</b>	<b>-0.172015</b>	<b>-1.890676</b>	<b>0.058668</b>
W*Population	-0.068599	-0.498982	0.617792
W*Gross State Product	0.065403	0.504635	0.613815
W*Wage	0.286868	1.269713	0.204187
<b>spat.aut.</b>	<b>0.163004</b>	<b>2.933298</b>	<b>0.003354</b>
<b>teta</b>	<b>0.198603</b>	<b>3.019479</b>	<b>0.002532</b>

### C.4.4 Reefer non-Managed

R-squared 0.4159; corr-squared 0.291; sigma<sup>2</sup>

0.0113; log-likelihood 180.63399

Variable	Asymptot	t-stat	z-probability
Intercept	0.390346	0.332283	0.739676
<b>Distance</b>	<b>-0.168329</b>	<b>-5.31797</b>	<b>0</b>
<b>Unemployment</b>	<b>-0.084566</b>	<b>-1.899403</b>	<b>0.057511</b>
Fuel	0.014454	0.181276	0.856151
Population	0.018397	0.167541	0.866945
Gross State Product	-0.027554	-0.261471	0.793729
Wage	-0.008364	-0.051433	0.958981
W*Distance	-0.009554	-0.340545	0.733446
W*Unemployment	-0.083775	-1.737777	0.08225
<b>W*Fuel</b>	<b>0.174115</b>	<b>2.517036</b>	<b>0.011835</b>
W*Population	0.1124	1.055891	0.291018
W*Gross State Product	-0.093342	-0.924521	0.355215
W*Wage	0.320909	1.657983	0.097321
<b>spat.aut.</b>	<b>0.28525</b>	<b>5.837377</b>	<b>0</b>
<b>teta</b>	<b>0.067494</b>	<b>1.634099</b>	<b>0.102238</b>