# Center for Excellence
# in
# Logistics and Distribution

National Science Foundation sponsored Industry/University Cooperative Research Center
**University of Arkansas/ University of Oklahoma/ Oklahoma State University**
**Texas Tech University/ Clemson University/ University of Missouri**
**UC Berkeley/ Virginia Tech/ Arizona State University**

## Final Research Report
**Project #: AZ10-METR**

## Development of Logistics Efficiency Metrics

**Industry Partner:** *CELDi (Center Designated Project)*

# Development of Logistics Efficiency Metrics

**Research Team:** J. Rene Villalobos, Arnold Maltz, A. Nicholas Mason
Cesar Meneses, Hector Flores, Liangjie Xue, Alfonso Ibañez

**Sponsor:** CELDi (CDP) | **Thrust Area:** Logistics

**Keywords:** Freight rate; Transportation Cost; Cost Index; Efficiency; Benchmarking

**Problem in Context:** The costs associated with product logistics represent one of the main drivers of a product's final cost. Thus, it is important to assess and monitor the performance of the logistics system to identify opportunities for improvement. This project focuses on developing a framework for tracking and improving the efficiency of logistics and transportation resources at an aggregate as well as individual scale. The developed framework yields two results: a logistics cost index to monitor system trends in addition to logistics performance and overall efficiency metrics to benchmark specific freight movement rates.

**Technical Approach:** Data was collected for all major U.S. lanes, which were segmented and analyzed to determine which factors directly affect the costs of transporting goods. Subsequently, the relevant factors identified were used to form a small set of clusters with the purpose of grouping lanes with similar behavior together. Ultimately, this small set of five clusters was used to summarize the behavior of the overall system and provide insights on the structure of the logistics network.

**Results:** The first result of the project is a global logistics cost index, developed as an aggregation of the rates on the five clusters. This index was validated through comparison with other well-known indicators of economic activity including the PMI, USIIP and CASS freight index. The analysis showed that the developed logistics cost index had valuable properties when compared to these indicators.The second result of this study is the analysis of a DEA based techniques for benchmarking purposes.

**Broader Value to CELDi Members:** The resulting index and corresponding efficiency rating metrics provide opportunities transportation cost reduction and as tools for the strategic sourcing of transportation services. For instance, the proposed index can be used for predicting future transportation rates, which will allow companies to effectively manage their resources. The proposed index also opens the possibility of creating a futures market for the trading of logistics related contracts.

**Future Research and Potential Extensions:** The framework developed can be expanded in future projects to obtain further information about the global logistics system dynamics.

As a potential expansion to the model it is desirable to consider much larger amounts of historical data for the creation and validation of an index for transportation and logistics costs. In this project, good results were obtained using historical data spanning 26 months. Nonetheless, using more historical data, a better estimate for the correlation of our index and other economic indicators can be achieved.

**2012**

# Development of Logistics Efficiency Metrics


J. Rene Villalobos, Arnold Maltz, A. Nicholas Mason

Cesar Meneses, Hector Flores, Liangjie Xue, Alfonso Ibañez

**Arizona State University   http://ilpil.asu.edu**



4/10/2012

# Table of Contents

# List of Figures

# List of Tables

## EXECUTIVE SUMMARY

Logistics plays a vital role in the performance of companies, particularly manufacturing companies, in many instances adding costs similar to or higher than the manufacturing cost of the final product. Commonly, the costs associated with logistics including warehousing, inventory and transportation average approximately 8% of the final price of the product (Davis and Drumm 1996). It is thus imperative to assess and monitor the performance of individual components in a logistics system to identify opportunities for improvement. Nonetheless, despite their importance, techniques for monitoring and controlling these costs are scarcely used in industry. This project focuses on developing a framework for tracking and improving the efficiency of logistics and transportation resources at an aggregate as well as individual scale; in particular, we focus on transportation outside of a specific company or supply chain, analyzing transportation throughout the entire continental US.

The objectives of the project include the development of an index to provide a periodic estimate of the efficiency of transportation/logistics resources as well as a set of metrics to facilitate the benchmarking of logistics operation outside of the company. Finally, the potential applications for a logistics cost index and the benchmarking capabilities for logistics resources are explored and a case for their further exploration is made.

Prior to developing a methodology for this study, we conduct a review of previous studies related to logistics performance measurement to identify what is already available to us, as well as to find any gaps within the field; the lack of research in the area of logistics metrics for large scale transportation is made apparent at this point. Upon the completion of this review, other indices both directly and indirectly related to logistics become available as useful resources. We then proceed to analyze these indices and determine whether any relationships existed amongst each other, which set the foundation for validating a logistics index in the future.

To develop our methodology, the project first focuses on identifying a single unit of measurement for the behavior of logistic costs and freight rates; the unit is defined as traffic lanes, which are a common unit of measure for all logistics companies. Data was collected for all major U.S. traffic lanes, a large volume of information which is reduced from 2465 to 680 data points through Pareto analysis. With a smaller sample size regression analysis is performed to determine which factors play a key role as performance indicators. These factors were found to be the travel distance as well as the origin and destination of these lanes, noting that trade imbalances also play a role but lack of quality in the data obtained made statistical significance hard to obtain. Finally, after identifying the significant factors that drive truck rates, we use them to form a set of five clusters of lanes. Clustering is made with the objective of grouping lanes that have a similar behavior, making their cluster a good summary for this subset of the network. Looking at five clusters we are able to simplify the analysis of the system and better understand its behavior, providing better insight on the structure of the logistics network.

Having summarized all lanes into clusters, we proceeded to create this single indicator for the overall system. A specific weight is assigned to each cluster to account for its relevance as measured by weighted distance, then a single global logistics cost index is created as a weighted average of all clusters. This single index is the validated through direct comparison with well established indices such as g the PMI, USIIP and CASS freight index. Our index was observed to have a strong correlation with the PMI; furthermore further decomposition and statistical analysis showed that amongst the 5 factors that compose the PMI, the components "new orders" and "inventories" yielded the best predictive model with an R-sq value of 83.2. This result is consistent with theory and intuition; furthermore, it provides exciting expectations for using the PMI as a potential predictor of logistics costs.

After developing the global logistics cost index, we proceeded to creating a metric that could be used to benchmark the performance of specific lanes. When looking at transportation on a large scale, as in this project, benchmarking can be difficult while a single index proves of limited use. Therefore, we decide to use data envelopment analysis (DEA) to rate lane's efficiencies. In our study the DEA model uses the attributes of lanes as inputs and the rate obtained from these attributes as the output, assigning an efficiency rating from 0 to 1 to each lane (1 being complete efficiency). The result is a flexible methodology for easily ranking the efficiency of logistics resources solely based on their characteristics as inputs and the rates obtained in each.

These seemingly simple results for a logistics index and its metrics counterpart can have significant implications in the way logistics resources are managed. For instance, we place special importance on the publishing of a single index which can be used for predicting future rates. In particular, the ability to make projections will allow companies to effectively manage their resources and can open the possibility of creating a futures market for the trading of logistics related contracts. What is more, further exploration for using the PMI as a predictor of logistics costs can aid logistics companies creating a significant impact though facilitating better planning and resource allocation.

This project proposes a framework for analyzing and utilizing information to achieve better logistics performance. However, the scope of our project was limited by data availability, which is one of the main factors preventing more research from occurring in this field. For this reason, a sharing platform for companies is proposed; this would allow for aggregation of information, as well as real time feedback on the efficiency of the lanes submitted. A framework for real time updating of the system status, immediate efficiency feedback based on DEA and a periodic updating and projection of a logistics cost index is detailed in the report. The consequences of implementing such a large scale information management system will be of great benefit to industry, allowing companies to better forecast, plan and negotiate contracts. Moreover, a better visibility of system status will also reduce the amount of empty travel incurred by logistics companies, improving their bottom line while minimizing their impact on the environment.

# 1 INTRODUCTION

The costs associated with logistics represent some of the main drivers of a product's final cost, where warehousing, inventory and transportation average approximately 8% of the final price (Davis & Drum, 1996). It is important to note that this cost is in many instances similar to or higher than the manufacturing cost of the final product. Furthermore, due to outsourcing, offshoring and other trends related to globalization, logistics costs are likely to rise in the future. Consequently, logistics will play an ever-expanding role in the overall performance of all businesses, in particular that of manufacturing companies. It is thus imperative to carefully assess and monitor the performance of individual components of the logistics system to identify opportunities for improvement and correction of these systems. Paradoxically, although logistics costs already represent a very important component of the product's final cost, techniques for the continuous monitoring and control of these costs are not very common in industry (Keebler, Manrodt, Durtsche, & Ledyard, 1999). This is even more surprising given the large amount of data currently available from enterprise execution and other real-time tracking systems.

The focus of this study is the development of a framework and specific tools for monitoring and improving the efficiency of logistics/transportation resources. Our aim is to identify the factors that determine the cost of transporting goods and use them to estimate costs of transportation, as well as to benchmark logistics operations. One specific feature of this study is that we will focus on logistics efficiency on a large scale, looking at the factors affecting efficiency without taking any company specific information into account; specifically, we will look at the general behavior of the cost of logistics resources throughout the entire continental US region.

The framework developed on this study will yield two related results: A logistics cost index to monitor system efficiency on a large scale and logistics performance metrics to benchmark the efficiency of specific instances of freight movement rates. Performance metrics should ideally reflect the efficient use of the transportation resources independently of the products transported as well as the geographical dispersion of the suppliers. These metrics will be widely applicable regardless of commodity or geographic area, making them amenable for benchmarking since two different companies can conceivably compare their operations even though their products and locations are dissimilar. Moreover, from the same framework, a global index is also developed which may be used in conjunction with the metrics developed. Therefore, under this framework, companies would be able to benchmark the usage of their resources against the rest of the system and to forecast their future costs more accurately.

Throughout this study, a methodology is developed to measure and benchmark the logistics performance of these networks. This measurement will be based on specific lanes, the main commodities transported through them and the different factors affecting their performance (both external and internal to the transportation networks). Once the most important factors affecting performance are identified, they can be used as a base for segmentation of lanes. This information can in turn be used to create representative indices and appropriate global metrics. The general methodology can be found in Figure 1-1 below.

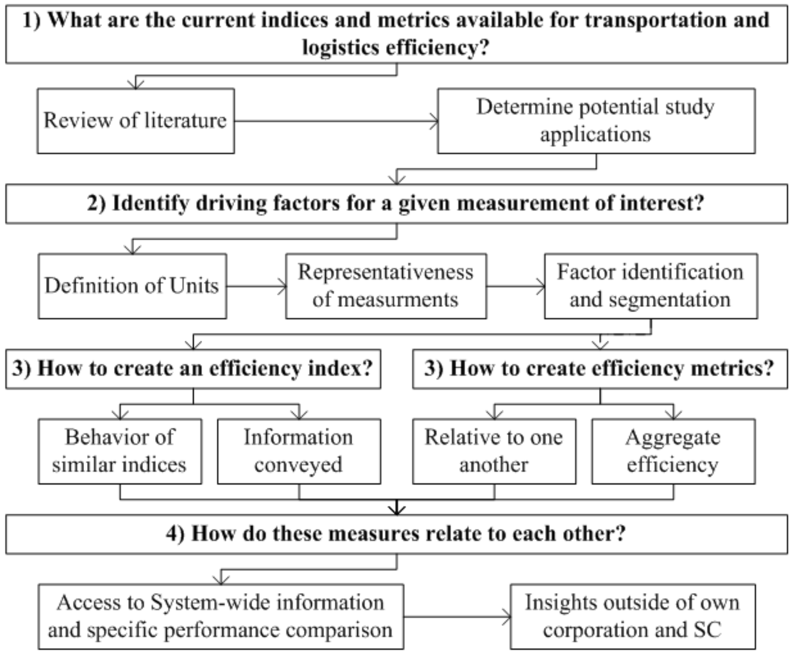


**Figure 1-1 – Overall Research Methodology**

Throughout the rest of this section we will expand on the composition of the methodology used and give an overview of the results from an application to a case study and potential applications.

## 1.1 Methodology Composition

In order to create appropriate indices and metrics, the first step taken was to review the literature on these topics and to compile a set of indices to analyze and review their behavior (Section 3). Subsequently, in Section 4 we proceed to determine the driving factors important for assessing the efficiency of logistics resources. This section is divided in four parts: We initially define the units to be measured and focus our analysis only in the units (transportation lanes) that are representative of the overall system through Pareto analysis. The second part of this section focuses in the identification of the factors relevant for the performance of these lanes through regression analysis. The third part focuses on using these relevant characteristics to cluster/segment the lanes into distinct groups that will be easily differentiated amongst each other through the use of data mining techniques. Finally, the last part of this section seeks to validate the difference between these groups.

Section 4 provides the base for analysis in subsequent sections, since it gives us information on what characteristics must be taken into account as important cost drivers, as well as the dynamics of various different clusters of lanes and their relative importance. Building upon the results of section 4, section 5 develops an index that summarizes the behavior of rates in each cluster and uses the weights assigned to each cluster for developing a single global indicator. This global index is then compared to other indices to validate its representativeness. Section 6 details the development of a framework to benchmark the relative efficiency of various lanes against the rest of the lanes through the use of data envelopment analysis (DEA). Finally, section 7 focuses on possible expansions and applications of the methodology developed throughout this study; in this section we detail opportunities for using a predictive logistics index to the use of individual metrics for each transportation lane. Section 8 provides a brief discussion on the results of the study.

# 2 PROJECT OBJECTIVES

The focus of this project is on logistics operations outside of an enterprise; namely, we do not consider the movement of goods within a factory or between commonly transited facilities. Instead, the focus of the project is on long haul shipping operations for containerized cargo since these costs are more variable and in many cases subject to market forces and outside of a company's control. We believe that companies could greatly benefit from a framework that allows them to monitor and improve the efficiency of their external logistics operations, either by having better capacity planning, contract writing or service provider selection.

***The objective of this research is the development of tools to provide a periodic estimate of the efficiency of transportation/logistics resources used for transporting goods as well as a set of metrics to facilitate the benchmarking of logistics operations outside of the company.***

This will be accomplished by developing a set of indices that reflect the behavior of rates for "representative truck lanes", together with metrics to benchmark the efficiency of specific lanes as compared to others with similar characteristics. These indices or metrics will allow users to predict and assess the cost of transportation for particular goods on particular lanes. To achieve the overall objective of this project, the following objectives must be met, which will be explained in more detail in Sections 3 through 6:

1. **Define target components of performance metrics**

   To carry out the analysis, we must define a target performance indicator, which can be any figure of interest such as cost, rate, capacity utilized or cost per ton-mile. Once the target metric is defined, the characteristics that influence this metric are identified and further analysis is made based on these main drivers of cost.

2. **Develop and validate global index**

   The objective of an index under the context of this study is to reflect the behavior of the system at a given point of time. Furthermore, if the behavior of the index and its interaction with other economic indicators are analyzed, we can quantify the expected changes of our index in future time periods.

3. **Metric Development**

   Define a framework in which inter-company logistics efficiency can be benchmarked in a manner that is not company specific, but can rather be used by many companies.

4. **Identify potential expansions to the use of logistics metrics and indices**

Relate the results of this study to industry needs and future applications that will benefit industry through the use of a logistics index and metrics.

# 3 REVIEW OF EXISTING INDICES AND METRICS

Before exploring the potential findings of this project, a review of previous studies related to logistics performance was conducted to identify what is already available to us and to also identify the gaps that remain within the topic. The studies summarized in this section provide an overview of previous works related to logistics performance. A list of current logistic performance metrics and indices is explained, which provides insight into how certain companies within the industry measure their logistics performance. In addition, some characteristics of the existing metrics and indices are identified that are of importance when developing new measures.

We provide an overview of the current status of research related to logistics performance in the next two sections. The metrics summarized in section 3.1 detail the existing metrics and review studies related to common logistics metrics. The indices summarized in section 3.2 give an overview of some of the common indices used currently in the industry as indicators of economic activity and of performance in specific segments of industry. A complete literature review and overview of the indices and metrics used in transportation and logistics can be found in Appendix A.

## 3.1 LOGISTICS PERFORMANCE METRICS

Studies with a scope beyond a company or a supply chain level are very limited; this is due to the fact that it is much easier to collect information for a single company as opposed to a country or region. Some of the more traditional performance measurements include, price, time and cost. Common metric performance indicators will be defined in this section.

1. Performance Indicators and Metrics

Some common performance metrics were analyzed by (Ahumada, 2003), who reviewed simple measurements usually taken by companies and assessed their statistical properties under various applications, these include:

- **$/(Ton–Miles):** Is defined as all billing associated with the shipments over the weight multiplied by the absolute miles
- **$/Ton:** Is defined as all billing associated with the shipments over the weight of the shipment
- **Ton/Trip:** Total cubic weight of all shipments over the number of trips
- **$/(CubicTon-Miles):** Is defined as all billing associated with the shipments over the cubic weight multiplied by the absolute miles
- **$/(CubicTon-Real Miles):** metric is defined as all billing associated with the shipments over the cubic weight multiplied by the real miles

- **$/(CubicTon):** Defined as all billing associated with the shipments over the cubic weight

In this list we find the measure of "CubicTon;" this is nothing more than a conversion of units using the density of the shipment in case that the main restriction is volume rather than weight.

2. Supply Chain Operations Reference Model (Huan, Sheoran, & Wang, 2004)

   The SCOR Model was developed to describe and improve all the business activities associated with the different phases of procurement across several companies, necessary to satisfy customer's demand. The model provides standard process definitions, terminology, and metrics that enable companies to benchmark themselves against competitors and other companies.

Most definitions of "logistics performance" have a limit of vision, focusing mostly on specific companies and supply chains. However, a metric that can be applied more broadly, outside of the context of a specific company, is necessary and requires more insight into the factors that formulate a logistics metric.

## 3.2 Review of Indices

By reviewing indices commonly used, we identify factors that are of concern within the industry. The characteristics and information in these indices may potentially be useful to our study. The first group of indices presented in section 3.2.1 is used to measure transportation, including maritime and terrestrial transportation indices. The second group, presented in section 3.2.2, is not directly related to transportation, but provides additional sources of information and may potentially be useful in validating a transportation index.

### 3.2.1 Review of Transportation related Indices

The first set of indices presented is for application within the transportation industry. Indices such as the Baltic Dry Index and the Shanghai Containerized Freight Index aid in the process of contract writing and evaluation. The CASS Freight and Transportation Service Indices provide a gauge of current movement within the trucking industry.

1. Baltic Dry Index (The Baltic Exchange, 2010)

   The Baltic Dry Index is a daily average of prices to ship raw materials. It represents the cost paid by an end customer to have a shipping company transport raw materials across seas on the Baltic Exchange, the global marketplace for brokering shipping contracts. It measures the demand to move raw materials and precursors to production.

2. Shanghai Containerized Freight Index (SCFI) (Shanghai Shipping Exchange, 2009)

The SCFI is published by Shanghai Shipping Exchange. It shows the ocean freight and surcharges of individual shipping routes on the spot market. 15 individual shipping routes are included, which are worldwide, not only limited to those related to China. The index is updated weekly.

3. Logistics Performance Index – LPI (The World Bank, 2010)

   This index is the result of the aforementioned study supported by World Bank. It evaluates the logistics efficiency of each country. Unlike the previous two indices, the LPI is not a single index but a group of indices give measurement in different dimensions. The LPI is regularly updated, and is available from World Bank, with the latest publication made in 2010. The detailed evaluation methods can be found in the study from (Hausman, et. al, 2005)

4. CASS Freight Index (Cass Information Systems, Inc, 2010)

   This is a monthly Volume Index of Freight Expenditures and Shipments that is built upon information from transportation dollars and shipments of Cass Information System's clients. The company CASS processes over $17.5 billion in annual freight payables. The purpose is this index is to compare levels of shipment activity on a month to month basis.

5. Transportation Service Index (TSI) (Bureau of Transportation Statistics, 2010)

   This index is created by the U.S. Department of Transportation (DOT), Bureau of Transportation Statistics (BTS), measures the movement of freight and passengers. The index, which is seasonally adjusted, combines available data on freight traffic, as well as passenger travel that have been weighted to yield a monthly measure of transportation service output. The main transportation elements upon which the index is based on are: rail (passenger and freight), pipeline (petroleum and natural gas), transit, waterborne, trucking, and aviation (passenger and freight).

### 3.2.2 Additional Indices

The Purchasing Managers and the UCLA Pulse of Commerce Indices provide the user with a gauge of current economic situations based on indicators derived from the transportation industry.

1. Purchasing Managers Index (PMI) (Institute for Supply Management, 2010)

   The PMI is an indicator of the economic health of the manufacturing sector. It is issued by the Institute for Supply Management (ISM), which is a non-profit group boasting more than 40,000 members engaged in the supply management and purchasing professions. An explanation of the calculation of PMI can be found from

INVESTOPEDIA (Ryan Barnes, 2010). The PMI is a composite index of five “sub-indicators”, which are extracted through surveys to more than 400 purchasing managers from around the country, chosen for their geographic and industry diversification benefits. The five sub-indexes are given by the PMI a weighting, as follows:

- Production level (.25)
- New orders (.30)
- Supplier deliveries (.15)
- Inventories (.10)
- Employment level (.20)

A diffusion process is done to the survey answers, which come in only three options; managers can either respond with "better", "same", or "worse" to the questions about the industry as they see it. The resulting PMI figure (which can be from 0 to 100) is calculated by taking the percentage of respondents that reported better conditions than the previous month and adding to that total half of the percentage of respondents that reported no change in conditions.

2. UCLA Pulse of Commerce Index (PCI) (UCLA Anderson School of Management, 2010)

   The Ceridian-UCLA Pulse of Commerce Index by UCLA Anderson School of Management is based on real-time fuel consumption data for over the road trucking and serves as an indicator of the current state and possible future direction of the U.S. economy. By tracking the volume and location of diesel fuel being purchased, the index closely monitors the over the road movement of produce, raw materials, goods-in-process and finished goods to U.S. factories, retailers and consumers.

3. USIIP

   The Federal Reserve's monthly index of industrial production and the related capacity indexes and capacity utilization rates cover manufacturing, mining, and electric and gas utilities. The industrial sector, together with construction, accounts for the bulk of the variation in national output over the course of the business cycle. The industrial detail provided by these measures helps illuminate structural developments in the economy. The production index measures real output and is expressed as a percentage of real output in a base year.

## 3.3 INDEX CORRELATION

Upon review of the various indices existing in different segments of the economy, we proceed to show the interaction between these indices. The first step in the assessment of these indices is estimating the amount of correlation that exists between them; this serves as an aid to develop and validate our index. In Figure 3-1 one can appreciate the relative degree of correlation that exists between the indices explained previously. A straight red diagonal means a direct 1:1 relationship between the factors. As one can observe, there exists a slight correlation between the CASS Freight shipment and the PMI Inventories index. Furthermore, there seems to be a relative high correlation between Fuel prices (U.S. Energy Information Administration, 2012) and the CASS Freight Expenditure and US Index of Industrial Production.

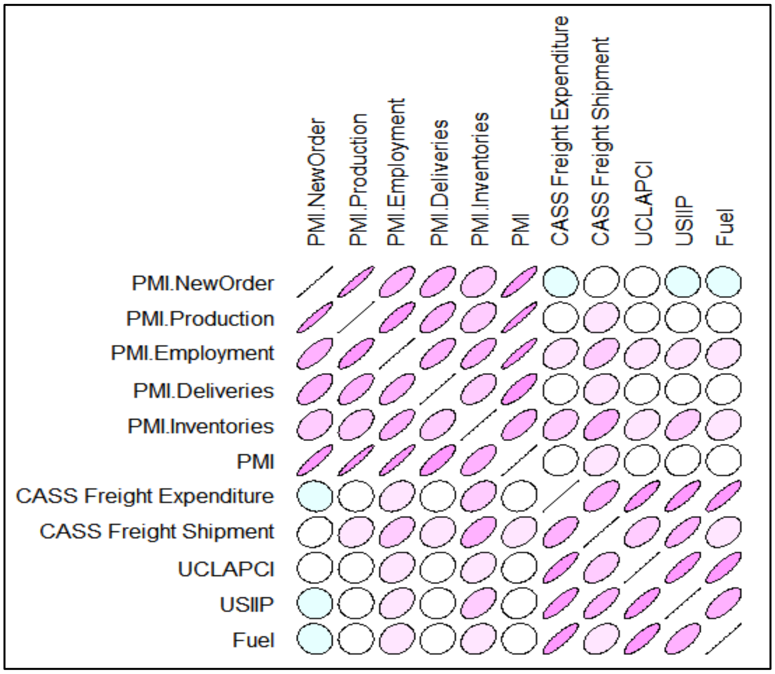


**Figure 3-1– Correlation between Indices**

In Figure 3-2 one can observe a correlation lag of approximately 6 months between the USIIP and the PMI. This is important since it shows that seemingly uncorrelated indices may in fact hold similar information of the same economic factors when time lags are accounted for. This is in essence one of the objectives of this project; being able to capture the behavior of economic movement in index form and applying the information related by this index in an application in the trucking industry.

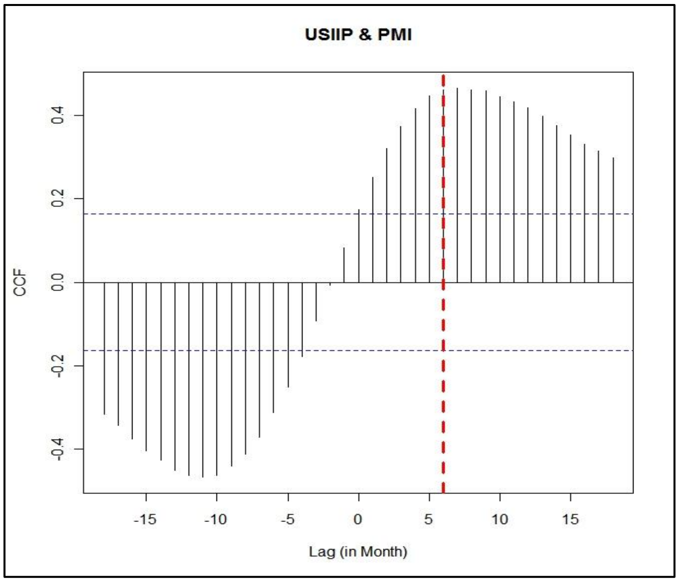


**Figure 3-2 – Cross Correlation between PMI and USIIP**

One of the ways in which the index developed by this project will be validated is using its correlation against indices within the economic sectors. Given that our index should capture the behavior of similar factors, it should behave in similar fashion as the PMI and USIIP. Thus, once our index is developed, a correlation analysis will be performed against these indices for validation purposes.

## 3.4 VOID IN CURRENT INDICES AND METRICS

There are a very limited amount of studies with a scope beyond a particular company or supply chain level. By reviewing previous studies it is shown that there is no standard definition for "Logistics Performance," it is completely dependent of the context for which it is applied. Each study has different research methods that were applied and differ greatly from each other. Furthermore, due to logistics efficiency being such a vital part of a company's performance, there is a great need for logistics performance measurements that aid in the operations of a company. Such measurements will aid in decisions pertaining inventory storage, movement of goods and contract negotiations. Since the current project aims to evaluate logistics performance of a lane, our results may be adopted by multiple participants using the lane, as opposed to just one particular company.

## 3.5 AVAILABILITY OF DATA INFORMATION RESOURCES

In order to determine what factors are most important to include in the indices, we determine what information is available. We use information provided by TransCore (TransCore, 2010) which provides current spot market and contract rates for the lanes used. From this information we can see what the current lane rates are, and then use historical data for the past years to assess how the rates change from month to month. This information is purchased at standard access rate. One can also obtain information for 13 months of historical data from the website's user interface.

Before continuing with the project, it must be emphasized that the acquisition of data through TransCore is limited to spot and contract rates. Since these are the only measures of efficiency that can be obtained, the methodology and results in sections 4 through 7 are based entirely in spot rates as a measure of the overall system efficiency. Likewise, metrics are developed taking rates as the target. Nonetheless, the methodology developed throughout the project can be expanded to other measures of efficiency such as $/ton-mile if data were to become available for validation with a different dependent variable.

# 4 ANALYSIS OF DRIVING FACTORS

In order to develop logistics indices and metrics, it is necessary to properly define what will be measured, as well as an appropriate and representative measurement unit. The right choice of units is a critical part of the project since ultimately the rest of the project will build upon the units that we define. The following sections describe the process followed for defining units and for performing further analysis: Section 4.1 specifies the definition of units; Section 4.2 through 4.4 specify the step by step analysis performed on the selected units and available information.

## 4.1 DEFINE A MEASURABLE UNIT

In order to define the right units to be measured and the target efficiency metric, we must first identify three important characteristics: what entities, instances or processes we need to compare, what the characteristics of the measured entities are and what the target metric is. For the scope of this project, we compare logistics efficiency in a macro-level outside of a company's range; therefore, routes transited throughout the US would be the logical choice. The characteristics of these routes and the target measurement unit are then defined based on the objectives of the project.

A key outcome required for developing a logistics metric that is widely applicable to companies is defining a measureable unit that is independent of company-specific terms and goods. Based on data from the Federal Highway Administration and the Economic Census, the unit has been defined as **a traffic lane**. A traffic lane is defined as the shortest physical route connecting any of two main statistical areas in the United States, and because a lane is the most impartial factor, using lanes as the unit of measurement allows the most flexibility for users when comparing different instances of transportation. A lane is defined by the following parameters:

- Statistical area of origin
- Statistical area of destination
- Distance (practical miles and Euclidean distance
- Direction (north/south, east/west)
- PADD region (Petroleum Administration and Defense District).

The statistical areas and PADD regions in the United States are illustrated below in Figure 4-1 and Figure 4-2 respectively.

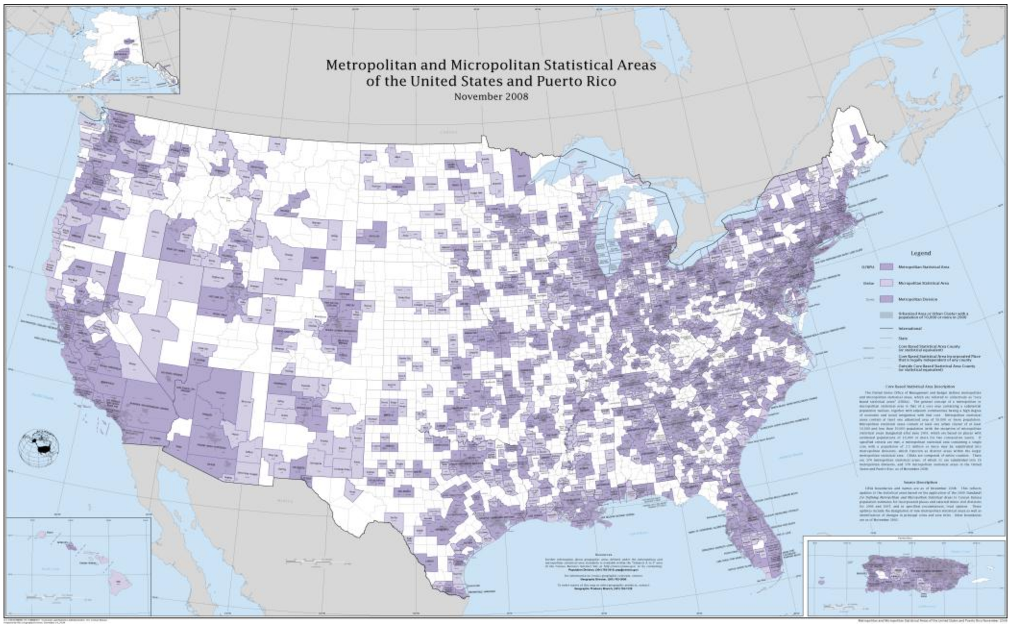


**Figure 4-1 – Statistical Areas in the United States (U.S. Census Bureau, 2008)**

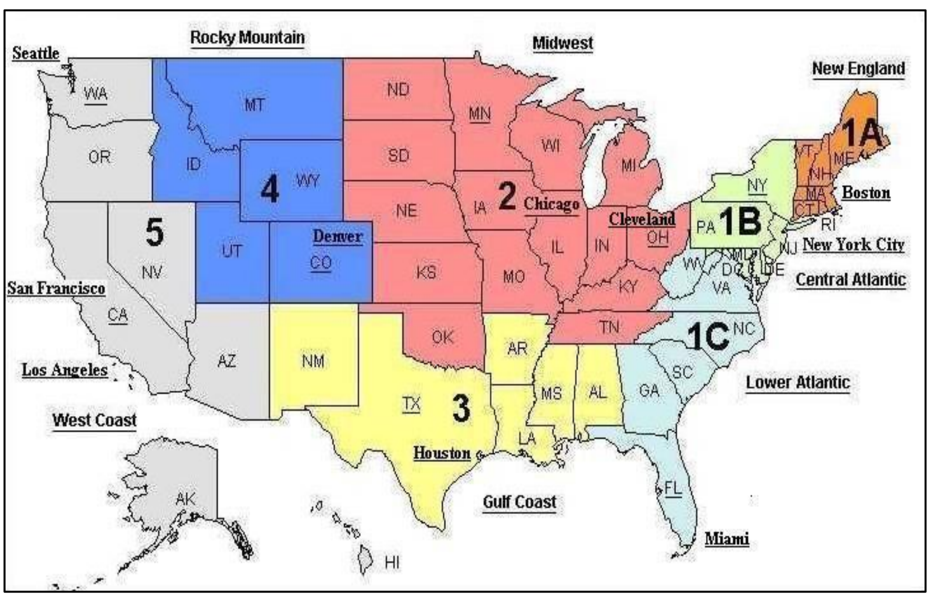


**Figure 4-2 – PADD Regions (U.S. Energy Information Administration, 2011)**

The use of statistical areas and PADD regions will give us enough flexibility to retrieve information specific to each origin and destination region, but it will also allow us to aggregate the statistical areas into more manageable and representative characteristics to represent each lane.

## 4.2 SEGMENTATION FOR ANALYSIS

Upon definition of the units of measurement it was found that the large volume of lanes that needed to be analyzed would make appropriate statistical analysis difficult to perform. Moreover, it was believed that a smaller and representative sample size could yield results of the same quality as the whole dataset. In order to get an appropriate representative sample of all of the lanes being modeled, analysis was done on the lanes by several modes to determine how the lanes would be segmented. The lanes were identified and analyzed by frequency of use, commodities transported, mode of transportation, and Pareto classification. The information for the rates of representative lanes was acquired using TransCore (TransCore, 2010) data.

The first part of analysis consisted of reducing the number of lanes considered in order to identify those with higher representative value. A Pareto analysis was performed in order to determine the top 20% of lanes which carried 80% of the weighted distance (measured by ton-miles). The ranking of the lanes was based on weighted distance to approximate the relative economic importance of the lanes; moreover, the weight transported through each lane was calculated by selecting the subset of commodities that are transported using standardized containers. A list of the commodities used to calculate the total weight transported in each lane can be found in Appendix B. These lanes in essence represent the most important lanes in the US in terms of product movement.

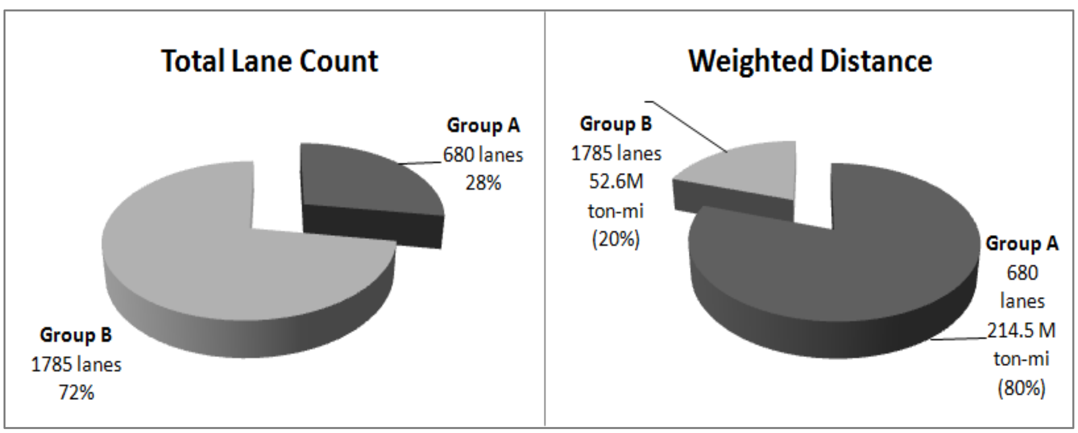


**Figure 4-3 - Pareto Lanes: Group A&B**

Figure 4-3 provides the results of the Pareto analysis on the lanes, demonstrating that 28% of the lanes move 80% of the total weighted distance. These lanes account for approximately 214.5 million ton-miles of the total 267.1 million moved in the US. This list of lanes is denominated as Group A and approximately contains around 680 lanes out of the total 2465 lanes analyzed; the rest of the lanes are contained in Group B. Group A lanes will be used primarily for analysis, while Group B lanes will be used for validation and testing purposes.

To test the validity of this segmentation, we expand the breakdown of group A and B lanes by heading, origin and destination. Figure 4-4 and Figure 4-5 present the segmentation of the lanes depending on their general heading. Likewise, Figure 4-6 and Figure 4-7 display the breakdown of lane count and weighted distance according to the lanes' region of origin and destination. Note that the Pareto proportion of 80-20% still holds when lanes are segmented according to general heading and PADD regions.

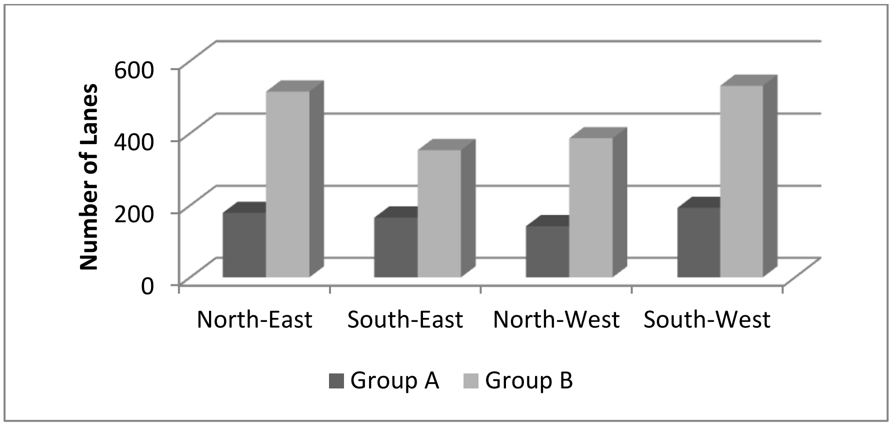


**Figure 4-4 – Pareto Analysis Results: Count by Lane Heading**

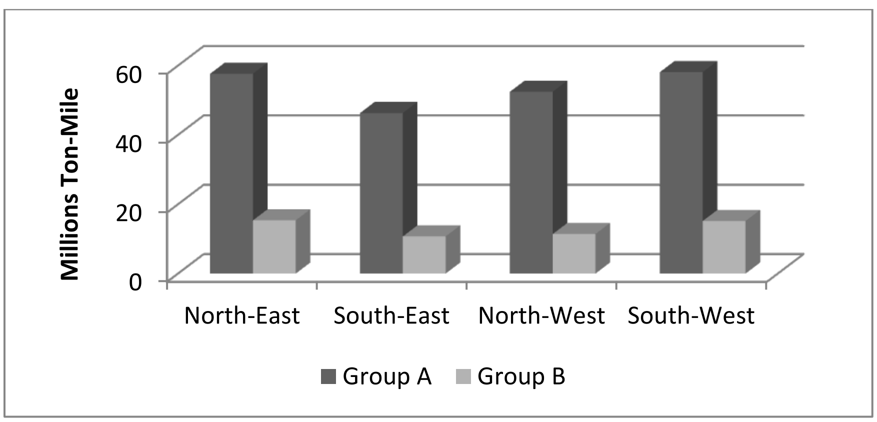


**Figure 4-5– Pareto Analysis Results: Weighted Distance by Lane Heading**

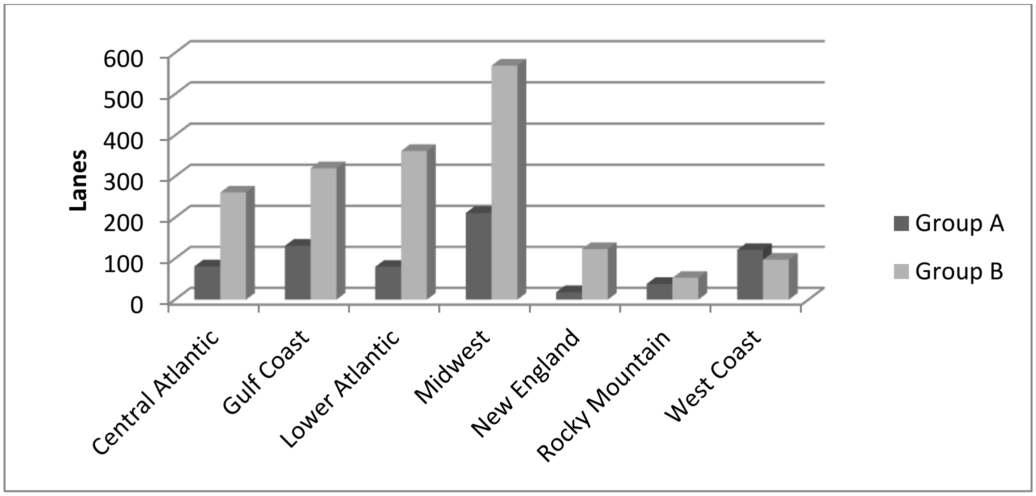


**Figure 4-6 – Pareto Analysis Results: Origin Regions**

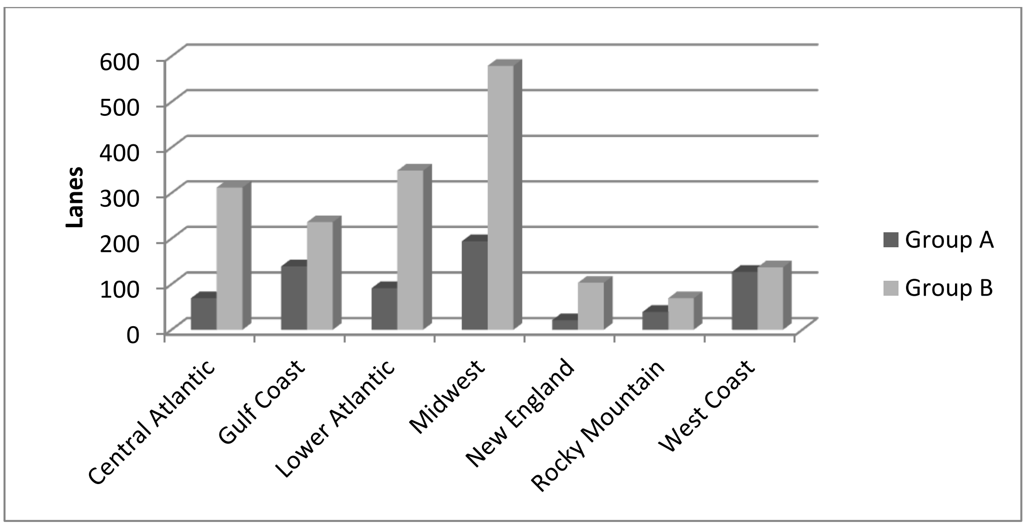


**Figure 4-7 – Pareto Analysis Results: Destination Regions**

For the purposes of this project, the rest of the analysis is based on Group A lanes, since these represent those with higher movement of goods based on the Pareto analyses performed. The lanes in Group B are kept for testing purposes in later analysis.

## 4.3 REGRESSION ANALYSIS

In order to measure logistics efficiency, we must first identify those factors which are relevant indicators of performance. In order to do this, we assume that the rates or costs in a particular lane can be predicted based on a limited set of variables representing the characteristics of that lane. The specific characteristics which are relevant can be determined through regression analysis; this is because those variables that add value to the predictive capabilities of the model are automatically selected through this method.

In order to define the factors which are relevant, a large set of characteristics pertaining to the cities of origin and destination is considered in the original models. Some of the characteristics considered in the model are shown in Table 4.1 , which summarizes a set of factors that are potential candidates for a good model. A full list of the economic factors considered in the model is found in Appendix C. All of the factors were considered as variables for both the origin and the destination by creating the respective variables for each instance.

**Table 4.1 – Possible Factors**

| **Suggested Responses** | **Suggested Factors:** | **Details** |
|---|---|---|
| Rates | Distances: | Practical Miles |
| Cost (Rate*Distance) | Origin/Destination: | Geographical regions<br>Lane heading |
| | Economic Factors: | Weight per year, commercial ports in the region, transportation companies, warehousing, road congestion |

A regression model was created using the characteristics of all lanes as a set of regressors as well as the appropriate transformations. The interaction between distance and spot average; and between distance and total cost is depicted in Figure 4-8. As one can observe there is a near linear relationship between the distance and total cost; however, a transformation would be required if only considering rates. Through regression analysis, irrelevant factors are removed based on their statistical significance to the regression model. Based on the examination of the standard summary statistics, such as the t and F statistics, or $R^2$, a regression model of considerable adequacy is selected. To further ensure model adequacy, basic residual analysis is performed to ensure that the basic regression assumptions are not violated.

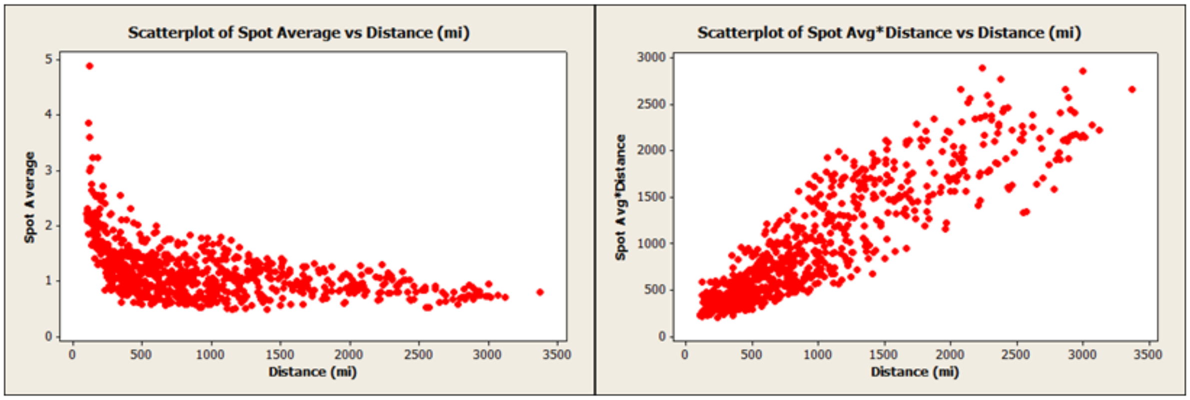


**Figure 4-8 – (a) Spot average vs. distance, (b) Cost (rate*distance) vs. distance**

Based on the regression analyses, it was determined that distances and origins/destinations (as indicator variables) are in fact relevant factors when understanding the behavior of lane spot rates in the US. These factors generated adequate models for predicting spot rates. On the other hand, none of the economic factors considered are relevant in the predictive model and do not explain the variability in the spot rates. These factors are discarded from further use.

However, we must emphasize the use of one particular economic indicator which is that of trade deficit/surplus at the origin/destination. When performing the regression with the geographic region parameters, imbalances were found to have no significance; nonetheless, in absence of these, imbalances played a role in predicting the rates. We mention this variable in particular because the information used to calculate trade imbalances was an aggregate of year 2007. Given that the analysis is made for a particular month, trade imbalances during that specific month would be more appropriate; unfortunately, no such data sources are currently available publicly. It is believed, based on preliminary results, that with a greater level of granularity, trade imbalances would be a valuable addition to explore in the model.

The structure of the final regression model and the findings of the relevant parameters are later used in additional developments of the project. Table 4.2 presents the statistical properties of a finalized version of the regression model for a one month period (November 2010).

**Table 4.2 – Model Properties**

| Model Characteristics | |
|---|---|
| R-Sq | 93.3% |
| R-Sq(adj) | 93.2% |
| R-Sq(pred) | 93.1% |

Table 4.3 presents the characteristics of the finalized model including coefficient for each regressor, its standard error, t-statistic, p-values and variance inflation factors. In this table, the letter “O” stands for region of origin, while the letter “D” stands for region of destination. The transformed response variable in this case is Sqrt(Dist*Spot Rate). The spot rates used for this model were those for the month of November 2010. As one can observe, all the variables in this table have low P-values and low multicollinearity as shown by the variance inflation factors.

**Table 4.3 – Final Model Variables**

| Predictor | Coef | SE Coef | T | P | VIF |
|---|---|---|---|---|---|
| Constant | 16.2369 | 0.4804 | 33.80 | 0.00 | |
| DistPract_Mi | 0.02107 | 0.000592 | 35.57 | 0.00 | 15.19 |
| Dist^2 | -2.76E-06 | 2.1E-07 | -12.82 | 0.00 | 16.05 |
| Gulf Coast_O | 2.7258 | 0.4515 | 6.04 | 0.00 | 2.062 |
| Lower Atlantic_O | 1.3338 | 0.5033 | 2.65 | 0.008 | 1.71 |
| Midwest_O | 6.5393 | 0.4177 | 15.66 | 0.00 | 2.428 |
| Rocky Mountain_O | 3.8708 | 0.6508 | 5.95 | 0.00 | 1.453 |
| West Coast_O | 10.0402 | 0.4539 | 22.12 | 0.00 | 1.959 |
| Gulf Coast_D | -4.4231 | 0.4547 | -9.73 | 0.00 | 2.186 |
| Lower Atlantic_D | -2.1807 | 0.4988 | -4.37 | 0.00 | 1.876 |
| Midwest_D | -7.2643 | 0.4291 | -16.93 | 0.00 | 2.442 |
| Rocky Mountain_D | -3.8981 | 0.6535 | -5.97 | 0.00 | 1.501 |
| West Coast_D | -9.2501 | 0.4644 | -19.92 | 0.00 | 2.13 |

Figure 4-9 shows the results of the residual analysis performed on this model. The normal probability and residual plots suggest that the basic regression assumptions were not violated. This further confirms that the identified factors do in fact have a high impact on the variability observed in the spot rates.

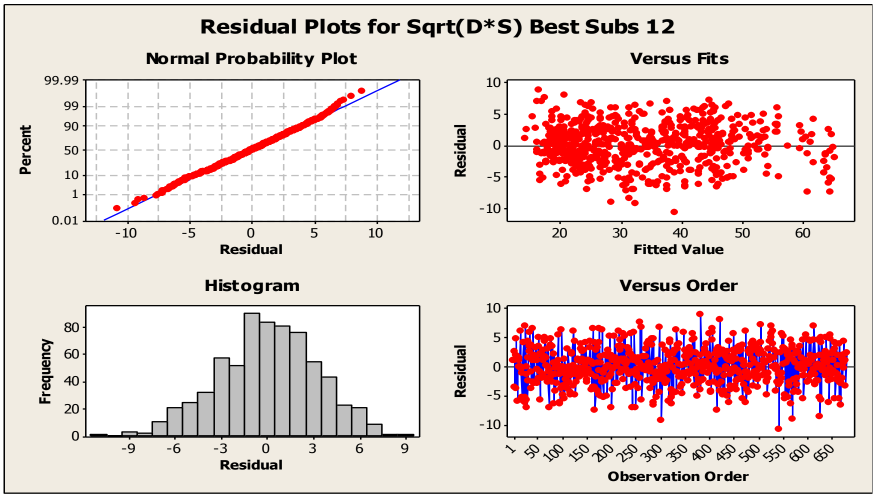


**Figure 4-9 – Residual Plots for Sqrt(D*S) Best Subs 12**

### 4.3.1 VALIDATION

In order to validate the model, the lanes contained in Group B (from the Pareto analysis) were used a testing set. By calculating the MSE for the Group B lanes and Group A lanes, it was found that the errors were not significantly different. Analysis showed an MSE of 10.9 and 10.74 for the group A and B lanes respectively. Since the errors of both models are not significantly different, we therefore confirm the validity of the model in addition to its applicability to other lanes in within the US.

### 4.3.2 REMARKS ON MODEL FOR FUTURE PERIODS

The adequacy of this model to accurately predict spot rates during other months of the year was analyzed. Not surprisingly, other months cannot be predicted based solely on the factors identified throughout the last section since the errors of prediction grow large as we move away from the month for which the model is made. Potential expansion to the prediction model could include time series-analysis and other seasonal variables such as movement of goods at particular months.

## 4.4 LANE CLUSTERING

Having determined what the significant factors driving truck rates are, we proceed to create a set of clusters for transportation. The purposes for clustering the transportation lanes are two: To create models that predict the rate behavior of lanes most accurately and to obtain better insights into the structure of the logistics network. When creating clusters, it is expected that the variability of rates within subgroups will be better explained by models made specifically for those sub groups. What is more, proper clustering of the lanes may reveal patterns that were hidden before in terms of the similarities within a group of lanes.

A simple, K-means algorithm with Euclidean distance measures and normalized data is used to cluster the lanes according to the characteristics found by the regression model. In total, 5 clusters are formed based on a cross-validation process using the factor suggested by the regression models. The 5 clusters were formed based on the following characteristics:

- Total movement (Kilo-Tons shipped from origin to destination)
- Distance
- Region origin and destination (as indicator variables)

The actual breakdown of the total lane population can be observed in Figure 4-10, which presents the percentage of lanes assigned per cluster. As one observes the distribution of the lanes is relatively well spread among the five clusters created.

### 4.4.1 CLUSTER DESCRIPTION AND INTERPRETATION

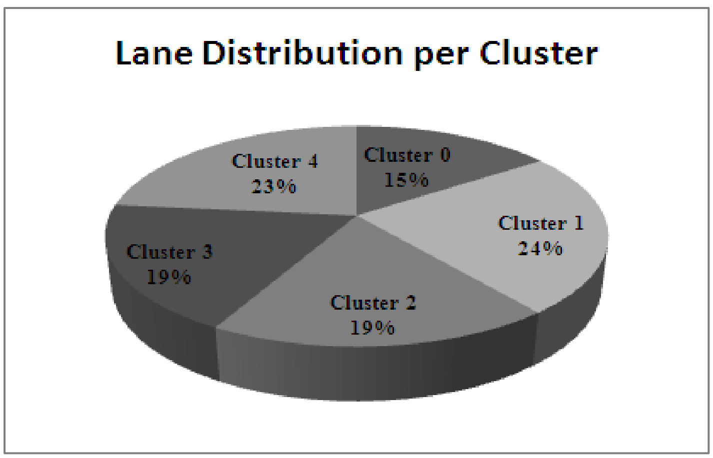


**Figure 4-10 – Lane Distribution per Cluster**

The characteristics of each cluster are further summarized in Table 4.4; for each cluster, the total number of lanes is presented as well their percentage share. The minimum, maximum, and average for the lane distance (in practical miles), as well as for the weight movement (in kilo-tons), are also given. Table 4.5 presents the counts of all the indicator variables for each origin and destination. This table color-codes each field depending on the relative ratio of the total lanes that have that specific characteristic within a cluster. A strong contrast between colors indicates that a cluster contains a majority of lanes with a specific origin or destination; whereas if contrast is lower the spread of characteristics within the cluster is more evenly distributed. Using these two tables we can infer some of the lane characteristics such as their relative importance and the general distribution of their geographical characteristics.

**Table 4.4 – Clusters Summary (Lanes, Distance, Weight)**

| Clusters | Lanes | | Distance | | | Weight | | |
|---|---|---|---|---|---|---|---|---|
| # | Count | % | Min | Max | Ave | Min | Max | Ave |
| 0 | 102 | 15% | 28 | 2500 | 806.9706 | 41.0241 | 10366.49 | 665.8072 |
| 1 | 158 | 24% | 54 | 2444 | 890.538 | 47.3035 | 5978.802 | 547.5254 |
| 2 | 124 | 19% | 81 | 2312 | 887.0161 | 38.082 | 3360.73 | 456.3847 |
| 3 | 124 | 19% | 90 | 3371 | 1743.113 | 29.8997 | 6082.505 | 536.3203 |
| 4 | 155 | 23% | 71 | 3002 | 1139.006 | 37.4201 | 3967.845 | 367.1429 |

**Table 4.5 – Clusters Summary (Origin and Destination)**

| Clusters | Origin | | | | | | Destination | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | New England | Central Atlantic | Lower Atlantic | Midwest | Gulf Coast | West Coast | New England | Central Atlantic | Lower Atlantic | Midwest | Gulf Coast | West Coast |
| 0 | 7 | 64 | 12 | 0 | 14 | 0 | 7 | 21 | 12 | 34 | 13 | 0 |
| 1 | 0 | 0 | 0 | 115 | 0 | 35 | 4 | 15 | 2 | 131 | 0 | 0 |
| 2 | 3 | 0 | 16 | 37 | 42 | 21 | 0 | 0 | 0 | 0 | 124 | 0 |
| 3 | 3 | 17 | 13 | 39 | 16 | 28 | 0 | 0 | 0 | 0 | 0 | 124 |
| 4 | 3 | 0 | 38 | 15 | 54 | 34 | 9 | 27 | 75 | 26 | 0 | 0 |

From the information in Table 4.5, we can assign a description to each of the groups obtained through the clustering analysis:

- Cluster 0: Origin Central Atlantic to balanced destinations —Short distance
- Cluster 1: Origin/destination Midwest
- Cluster 2: Balanced origins with destination Gulf Coast
- Cluster 3: Balanced origins with destination West Coast
- Cluster 4: Balanced origins with destination Lower Atlantic—Long distance

For Table 4.4 we can infer the relative importance of each cluster, as well as the dispersion amongst origins and destinations. For instance, although cluster 0 has fewer lanes and all regions will be on average closer, its relative importance may be greater since the weight transported is almost twice as much as cluster 4. This result is consistent with the descriptions and characteristics of the clusters such as cluster 0, which consists of locations that are mostly of large population density. A closer analysis of each cluster can be found in Appendix D.

### 4.4.2 CLUSTER REGRESSION ANALYSIS

After the clusters were formed a regression analysis was performed for each individual cluster in the same manner as done in section 4.3 for the whole population of lanes. The regression models developed for the lanes within each cluster are based on the following factors: geographic origin and destination, distance, and movement. These models adequately explain the variability of the quoted rates for lanes with particular characteristics. The models are formulated based on rates from November 2010. In this case, the response variable for the models is a transformation of the total transportation cost ((distance*rate)^0.5).

Before developing the regression models per cluster, pre-processing must be done on the data to eliminate any outlying data points that would significantly affect the models. From this pre-processing, we observe that lanes with distances less than 200 miles would not be considered because of their unusual characteristics. This is demonstrated by Figure 4-11. As one can observe from this figure, lane spot rates with a distance below 200 miles tend to behave erratically. This suggests that additional factors are needed for these of lanes. Therefore, the framework developed through this analysis is only applicable to long-hauls.

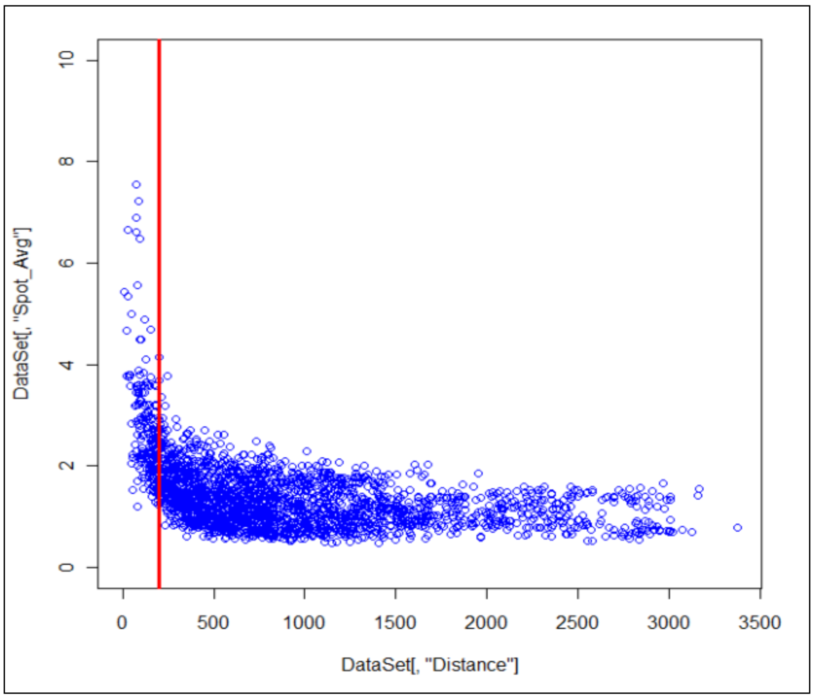


**Figure 4-11 – Identifying Uncharacteristic Lanes**

After pre-processing, a least square regression model was created for the lanes within each particular cluster. Table 4.6 provides the coefficient values for each of the regressors considered in the models.  As one can observe from this table, the configuration of the model was influenced by the composition of each cluster.

**Table 4.6– Coefficients per Cluster Model**

| | | Cluster Model | | | | |
|---|---|---|---|---|---|---|
| | **Variables** | **00** | **01** | **02** | **03** | **04** |
| | Transformation (^T) | 0.5 | 0.5 | 0.75 | 0.5 | 0.5 |
| | Constant | 19.46 | 20.13 | 112.84 | 20.67 | 25.15 |
| | DistPract | 0.01 | 0.03 | 0.11 | 0.01 | 0.014 |
| | Dist^2 | | | | | |
| **Origin** | New England | -4.44 | | | -9.88 | |
| | Central Atlantic | -4.84 | | | -8.13 | |
| | Lower Atlantic | -3.71 | | -74.86 | -7.88 | -6.42 |
| | Midwest | | -2.67 | -37.58 | | |
| | Rocky Mtn | -5.00 | -6.84 | | -2.38 | -5.403 |
| | Gulf Coast | | | -61.67 | -4.08 | -4.903 |
| | West Coast | | | | | |
| **Destination** | New England | 6.07 | 2.32 | | | 2.102 |
| | Central Atlantic | 5.39 | 3.78 | | | |
| | Lower Atlantic | 3.21 | -4.69 | | | |
| | Midwest | -2.68 | -4.15 | | | -3.985 |
| | Rocky Mtn | 2.97 | | | | |
| | Gulf Coast | | | | | |
| | West Coast | | | | | |

The adequacy of these models is gauged based on the model properties, such as the R-Sq (adj) and R-Sq (pred) values, which measure the level of variability captured by the regression model. In this case, we find that the models created for each of the clusters are adequate for explaining the variability in the spot rates and predicting new lane spot rates (Table 4.7). The regression model for cluster 01 has the highest values of R-sq (adj) and R-sq (pred) with values of 96.26% and 95.85%, respectively.

**Table 4.7 – Regression Models per Cluster**

| Cluster | R-sq (adj) | R-sq (pred) |
|---|---|---|
| **00** | 86.47% | 83.05% |
| **01** | 96.26% | 95.85% |
| **02** | 95.14% | 94.83% |
| **03** | 80.00% | 79.01% |
| **04** | 91.70% | 91.25% |

The regression models made for each cluster serve two purposes: validation of the clustering criteria and to provide a predictive model to account for the variation in rates for each specific cluster. Moreover, they also show that effective predictive models can be made on each cluster and applied in real life settings as suggested by the R-sq (pred) for each cluster. Those clusters which have a lower R-sq (pred) would also be candidates for considering additional factors or further segmentation.

Clustering the lanes will facilitate the analysis and interpretation of the behavior of specific groups of lanes. Having the ability to label a given lane as falling into a specific group makes analysis and future projections on their behavior more reliable and easily accessible. Moreover, clustering would allow for a more targeted analysis of the behavior of each group to avoid generalizing all lanes within the US as behaving identically.

### 4.4.3 REMARKS ON PREDICTIVE CAPABILITIES OF CLUSTER MODELS

To test the predictive capabilities of a single formula for each cluster, the regression models created in section 4.4.2 for the month of November 2010 were used for all other months from Sep-09 to Sep-10. The predicted rates for each lane were compared to the real value obtained and the mean square error (MSE) from the prediction to the actual value was computed for each cluster. The calculations for accuracy as measured by the MSE for each cluster can be seen in Table 4.8 below.

**Table 4.8 – R-sq (pred) per cluster through time**

| | Sep-09 | Oct-09 | Nov-09 | Dec-09 | Jan-10 | Feb-10 | Mar-10 | Apr-10 | May-10 | Jun-10 | Jul-10 | Aug-10 | Sep-10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cluster 0** | 84.3% | 94.6% | 95.4% | 96.3% | 96.2% | 95.3% | 94.6% | 93.1% | 90.5% | 93.7% | 93.1% | 92.2% | 93.0% |
| **Cluster 1** | 75.9% | 75.0% | 79.0% | 75.5% | 78.1% | 80.3% | 83.4% | 85.9% | 68.7% | 84.4% | 80.8% | 83.7% | 83.8% |
| **Cluster 2** | 44.0% | 48.5% | 34.6% | 51.2% | 57.7% | 62.3% | 71.6% | 67.9% | 64.6% | 69.6% | 66.8% | 70.0% | 74.2% |
| **Cluster 3** | 84.3% | 82.1% | 74.8% | 80.3% | 82.3% | 80.3% | 82.2% | 80.4% | 77.1% | 72.1% | 73.7% | 78.0% | 80.8% |
| **Cluster 4** | 65.8% | 72.4% | 66.8% | 59.8% | 68.5% | 65.7% | 70.0% | 61.4% | 63.5% | 66.8% | 71.2% | 69.9% | 75.3% |

When applied to different time periods the regression models made for each cluster, using Nov-2010 for building the model, are shown to have poor predictive capabilities with the exception of cluster 0 as indicated by Table 4.8. The results above show that the behavior of rates in each cluster is in fact dynamic, and a static model is inappropriate for explaining it. This indicates that for prediction purposes different models should be created for each month and cluster combinations as time progresses; alternatively, a single time series model that accounts for the relevant factors should be made.

# 5 INDEX DEVELOPMENT

Having identified the factors that are relevant for predicting transportation costs and having segmented the lanes by clusters, we proceed to analyze the interactions amongst clusters and the implications of interactions. Moreover, since the clusters encountered should be representative of the overall system performance, there is a way to summarize the behavior of the clusters to make a single indicator for logistics and transportation in the US. In order to summarize the behavior of each cluster into one single index, we take a simple average of the costs of transportation for the lanes in that cluster.

To analyze the behavior of clusters it is of greater interest to observe the relative variation of the cost rather than their real average. That is, to analyze whether there is a relationship amongst the various clusters, it is more useful to look at them on the same scale. Therefore, to standardize the units of each cluster and simply observe their variation, we take the ratio for the costs of transportation on each month and we divide them by a baseline month, which was selected to be the predicted transportation cost for the month of Nov-2010.

Since the transformation is a simple scale difference, then the index developed can be easily transformed back to determine the average transportation costs observed in a particular month. The value of the index for each month and each cluster will therefore be given by Formula 1 below. The values of the index for each cluster through time are shown in Figure 5-1. The cluster specific indexes can be updated in time intervals as small as needed; however, in our case we use month by month variation in costs since it is the most accessible and easy to analyze interval.

(1) $$\boldsymbol{Index_{i,j} = \frac{Actual\ Cost_i}{Predicted\ Cost_{(Nov.10),j}}\quad for\ all\ months\ i\ and\ all\ clusters\ j}$$

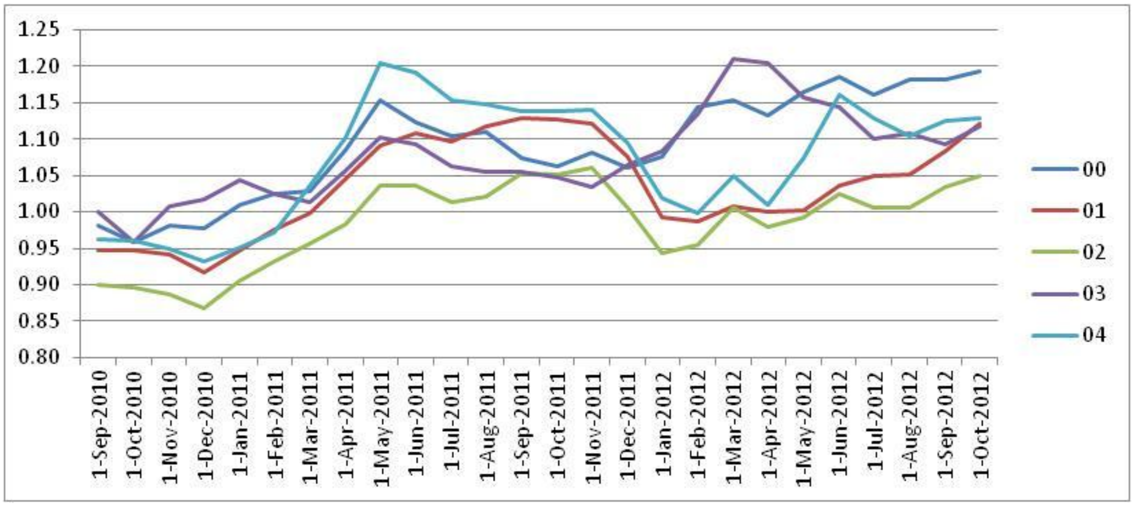


**Figure 5-1– Index values per cluster**

Now that we have defined an indicator that represents the behavior of each cluster, we proceed to analyze the relationship amongst the five clusters and seek to develop one single index that summarizes the behavior of all lanes.

## 5.1 RELATIONSHIP AMONGST CLUSTERS

We first explore the behavior of the five clusters compared to one another. That is, we seek to determine whether there is a significant correlation between clusters to know which lanes will experience changes in their cost that occur simultaneously. In Figure 5-1 we can observe that there may be a relationship amongst clusters; however, it is not readily apparent at a first glance. In Table 5.1 below we observe that some clusters show a much higher degree of correlation between each other. Mainly, we observe that clusters 01 (O/D. Midwest), 02 (D. Gulf Coast) and 04 (D. Lower Atlantic) are highly correlated. Likewise, the costs of transportation in clusters 00 (O. Central Atlantic) and 03 (D. West coast) are correlated.

**Table 5.1– Inter cluster variation**

| Cluster Correlation | 00 | 01 | 02 | 03 | 04 |
|---|---|---|---|---|---|
| 00 | 1.000 | | | | |
| 01 | 0.560 | 1.000 | | | |
| 02 | 0.737 | 0.946 | 1.000 | | |
| 03 | 0.822 | 0.207 | 0.454 | 1.000 | |
| 04 | 0.685 | 0.919 | 0.929 | 0.326 | 1.000 |

Since the indices for each cluster differ from each other significantly, special care will have to be taken when developing a global indicator.

## 5.2 SINGLE GLOBAL INDICATOR

In order to develop a single indicator that can be used as a reference for transportation costs, a decision is made to assign a specific weight for each cluster to account for its relative importance. The weights are given by the weighted distance of all lanes contained in each cluster, taking the relative share of the total weight for each cluster. Eqn. 1, used to calculate each weight, is given below. The relative weights of each cluster are found in Table 5.2.

$$Weight\ \%_{Cluster} = \frac{\sum_i Weight_i * Distance_i}{\sum_{Cluster}\sum_i Weight_i * Distance_i};\ where\ i\ \in Lanes \quad \textbf{Eqn. 1}$$

**Table 5.2– Relative weights per cluster**

| Cluster | Sum of W*D | Percentage |
|---|---|---|
| 00 | 26427078 | 13.355% |
| 01 | 44767203 | 22.623% |
| 02 | 28873009 | 14.591% |
| 03 | 57118796 | 28.865% |
| 04 | 40694224 | 20.565% |
| **Grand Total** | **197880309** | **100%** |

Using these weights, a single index is calculated. The behavior of this global index through time can be observed in Figure 5-2 below. This is a significant result that will be validated and further analyzed in the following sections.

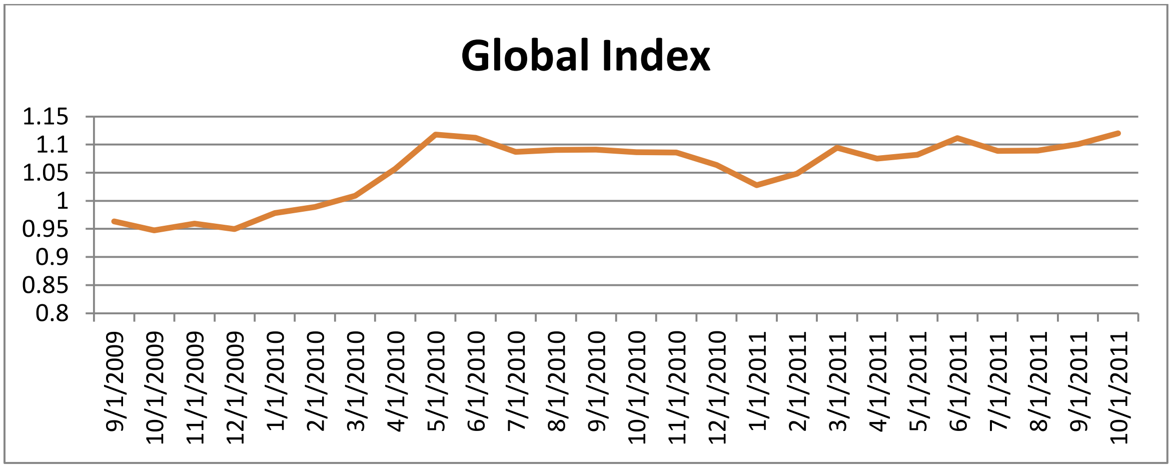


**Figure 5-2– Global transportation index**

## 5.3 INDEX VALIDATION

Note that even though an index representative of all clusters (and therefore all lanes) has been created, there is still a need to validate this index. In order to do this, our index is compared to other well known indices using the same methodology as the one used in section 3.3. The most representative indices to be used for the validation are thought to be: The PMI index, USIIP and the CASS freight index.

To validate this new transportation index, we calculated its correlation to the representative indices mentioned above. Furthermore, the correlation was also calculated through a series of time lags, moving our index forward and backward in time to test whether it had the potential to predict (or be predicted by) another available index. This potential is directly reflected by the degree of correlation achieved at a specific lag. The correlation of our index as compared to these three widely used indexes can be seen in Figure 5-3.

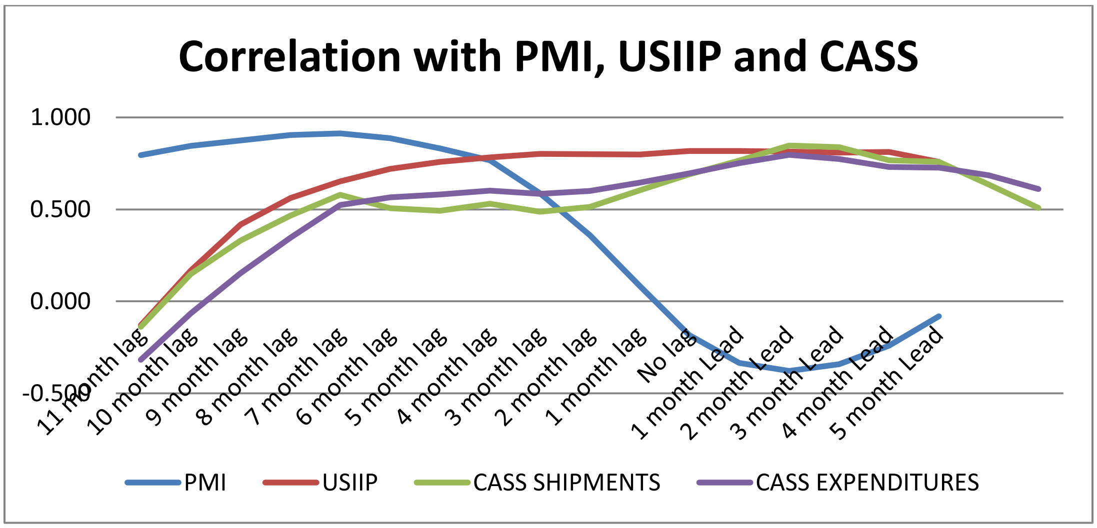


**Figure 5-3– Correlation with PMI, USIIP and CASS indexes**

Note that at first glance, our index is highly correlated with the CASS index, which has strong face-value because the CASS freight index measures specifically freight expenditures. The USIIP also has a strong correlation; however, the exact value is hard to determine because of the large spread that it has when accounting for lags. On the other hand, there is a significant correlation with the Purchasing Managers Index (PMI); what is more, the lag observed indicates that there may be a potential for using the PMI as a predictor for our index. This interesting finding is explored further in the following section, where the PMI is further decomposed into its respective components.

## 5.4 DEVELOPING A PREDICTIVE INDICATOR

After observing that our index had a strong correlation with the PMI, it was desirable to check whether the PMI could be used as a predictor of our index. Furthermore, since the PMI is composed of five factors (new orders, production, employment, deliveries and inventories), it was decided to test whether some specific components of the PMI could be used to predict our index to a better accuracy.

It was found, through regression analysis, that the components of "new orders" and "inventories" yielded the best predictive model for our index when accounting for a 6 month lag. These two components resulted superior to the aggregate PMI, yielding a predictive model with an R-sq(adj) of 83.2, as opposed to 82.5 for the aggregate PMI. This result further validates our index, since when dealing with transportation, the factors of the PMI that are most likely to have significant interactions with logistics costs are these two. A graph showing our index (actual and predicted) vs. the PMI is shown in Figure 5-4 below.

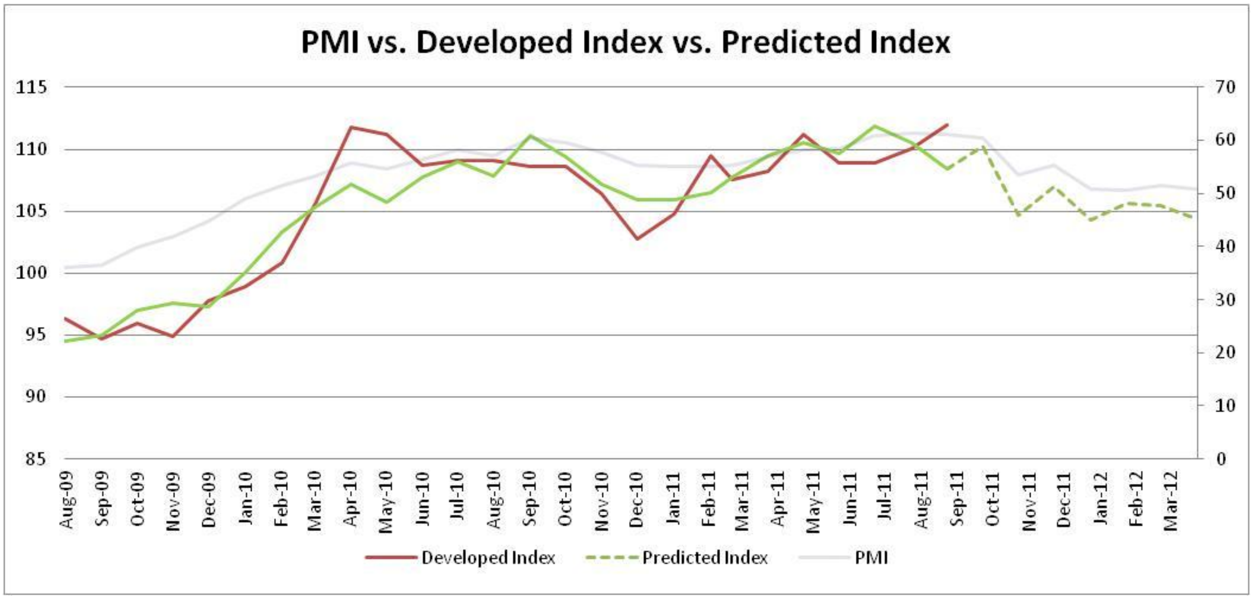


**Figure 5-4– PMI vs. (Developed and Predicted) Index**

From these results, it can be inferred that the PMI may have a good potential for predicting our index, which in turn would have strong implications of its own. More importantly, our index shows that it has strong validity as shown by its correlation with other economic indexes and it can also have strong potential as an indicator of economic performance and activity.

# 6 METRIC DEVELOPMENT

Developing an index is a valuable tool for estimating the state of the overall system and potentially making projections to the future. However, it has a strong void when dealing with the issue of benchmarking specific lanes or instances of contracts made for moving cargo along those lanes. It is because of this void that this section is dedicated to the development of metrics for assessing the efficiency of transportation resources on specific lanes.

When looking at transportation on a high level, as it is done in this project, benchmarking becomes more complicated since the points of comparison between companies are different and the amounts of data needed to make proper comparisons are also large. Moreover, logistics efficiency at the level of freight forwarding and long haul shipments is much harder to measure since the efficiency at this level is subject to market forces and the rates of transportation at the time that shipments are made. Because of the variable efficiency and the difficulties of comparison amongst different companies and different lanes, a new approach for comparison must be proposed.

Due to the difficulties of measuring absolute efficiency at this level, an approach for measuring the relative efficiency of a specific lane at a particular point in time is proposed instead. To do this, data envelopment analysis (DEA) is proposed as a potential solution. By measuring the relative efficiency of a lane or a specific shipment, it becomes easier to monitor its performance even if the overall rates of the system change. In this section we will describe some of the theoretical background behind DEA and the form in which it can be used to measure efficiency.

## 6.1 DATA ENVELOPMENT ANALYSIS FOR EFFICIENCY RATING

The main feature of DEA is its capability for rating a specific entity's efficiency on a scale of one to zero depending on its attributes and the attributes of similar entities being ranked. DEA uses the simple concept of an "efficient frontier" together with the fact that in order to obtain some output there has to be a set of inputs. The lanes in the efficient frontier are those which better utilize their inputs to obtain a given outcome; convex combinations of the inputs are combined in order to create an efficient/inefficient output (Molinero & Woracker, 1996). In the case of an "efficient" entity, the specific entity would be said to lie in the "efficient frontier" and would receive a rating of one. If an entity is ranked as "inefficient", it is because it fails to form part of the efficient frontier and there are other entities that lay in the efficient frontier whose convex combination could be used to construct the "inefficient entity". This concept is better illustrated by Figure 6-1.

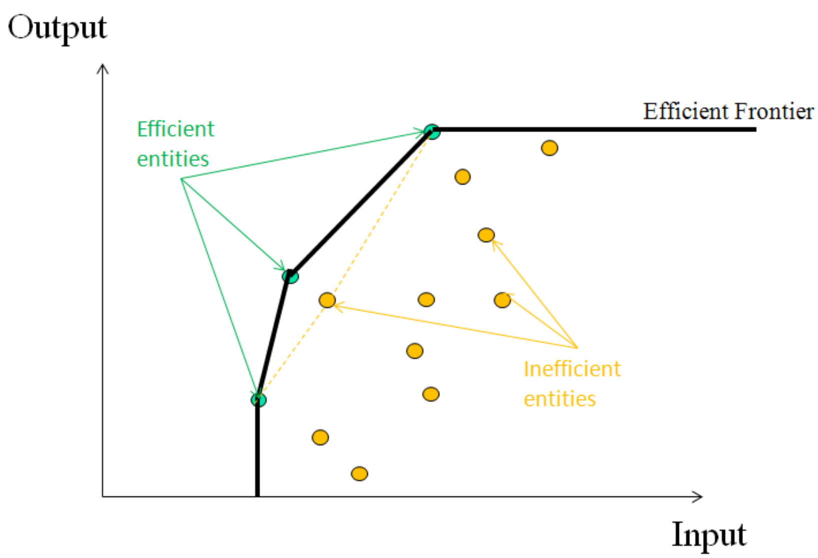


**Figure 6-1– Ranking of entities by DEA**

In the case of our study, the inputs to the DEA model would be the attributes of the lanes being considered, while the output would be the rate obtained for the lane (although other measures such as time and capacity utilization could be used). The concept of using DEA to rank the efficiency of lanes is illustrated by Figure 6-2 below.

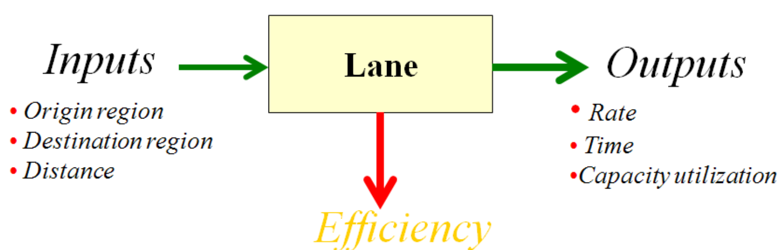


**Figure 6-2– Input/output for lane efficiency ranking**

### 6.1.1 MODEL FORMULATION

DEA is a problem formulated as a linear optimization program (LP), in which the efficiency is a decision variable that the model seeks to maximize by using the characteristics and information of all entities involved in the ranking (Charnes et. al, 1985). The formulation of the LP is shown below:

Min: $\theta_0$

Subject to:

$\sum_{i=1}^{n} \lambda_i x_{ij} \leq x_{oj}\theta_0 \quad \forall j$

$\sum_{i=1}^{n} \lambda_i y_{ik} \geq y_{ok} \quad \forall k$
$\lambda_i \geq 0 \quad \forall i$

Where:
$x_{ij}$ is level of input *j* for firm *i*
$y_{ik}$ is level of output *k* for firm *i*
$\theta_0$ is the efficiency of reference firm 0
$\lambda_i$ is the "intensity" of firm *i* (i.e. the amount of the firm used to make up virtual firm)
$n$ is the number of firms

One advantage of DEA is that being an LP model, the definition of the model and the solution procedure are easily obtained, even for large instances of data. Moreover, since in order to rank lanes efficiency the only information required are those lanes that are at the efficient frontier, data requirements can be further reduced and updated dynamically. A simple model was made for calculating the relative efficiency of all lanes using AMPL; the code for this model can be found in Appendix E.

### 6.1.2 APPLICATION OF DEA TO RANK THE EFFICIENCY OF LANES

To illustrate the use of DEA to rate the efficiency of various lanes, we use two sets of lanes with different characteristics. The first set of lanes, seen in Table 6.1, consists of lanes with the same origins and destinations, but different traveling distances. It can be seen from this table that three lanes form the efficient frontier; two others chosen to illustrate lanes that fall outside of the efficient frontier are given ratings of 0.92 and 0.88. These ratings can be replicated from a direct comparison with only those lanes in the efficient frontier. Note that instead of using rates directly, the reciprocal of the rates (miles/$) was used; this is because DEA relies on the "efficient" lanes being those with the highest values.

**Table 6.1– DEA ratings for subset 1**

| LANE | Distance | Origin Region | Destination Region | Rate ($) | Efficiency |
|---|---|---|---|---|---|
| L065121 | 2360 | MidwestO | West CoastO | 0.84 | 0.88095 |
| L065113 | 2291 | MidwestO | West CoastO | 0.82 | 0.91854 |
| L057113 | 2354 | MidwestO | West CoastO | 0.74 | 1 |
| L073121 | 1235 | MidwestO | West CoastO | 0.99 | 1 |
| L083113 | 1508 | MidwestO | West CoastO | 0.92 | 1 |

In the same manner, we illustrate the use of DEA for a different set of lanes with different origins and destinations on Table 6.2. In this case, we can see that the efficient frontier is now formed by five instead of three lanes; moreover, note that the "inefficient" lane is labeled accordingly because the rate paid in this lane is higher than that of the two surrounding lanes; that is, a "virtual lane" can be formed to have a better rate than lane L090095 through combining lanes L067094 and L090088 in a linear fashion.

**Table 6.2– DEA ratings for subset 2**

| LANE | Distance | Origin Region | Destination Region | Rate ($) | Efficiency |
|---|---|---|---|---|---|
| L052053 | 216 | MidwestO | Gulf CoastO | 1.54 | 1 |
| L090088 | 341 | MidwestO | Gulf CoastO | 1.25 | 1 |
| L090095 | 463 | MidwestO | Gulf CoastO | 1.3 | 0.94037 |
| L067094 | 761 | MidwestO | Gulf CoastO | 1.13 | 1 |
| L065095 | 1426 | MidwestO | Gulf CoastO | 0.98 | 1 |
| L065097 | 1667 | MidwestO | Gulf CoastO | 0.96 | 1 |

As it can be seen, DEA can consistently and predictably rank the efficiency of the rates for various lanes in a manner that is consistent with intuition. For any set of attributes, an unbiased ranking can be given to lanes. Moreover, since the analysis gives a relative comparison to other lanes, anonymity of information can be preserved by only showing the ratings, but not the actual rates paid.

### 6.1.3 Expansion of DEA

Having the information for all lanes, it is possible to calculate the efficiency of each individual lane at each time period. All of these efficiencies are then stored and compared to one another in the same manner in which we would compare costs or rates; however, we have the significant difference that we are now comparing numbers which are normalized according to the specific attributes of each lane.

If lanes with similar attributes are compared, then a change in the efficiency score of one single lane indicates that the costs in that lane have changed with relation to the others. Conversely, suppose that transportation rates increased by the exact same scale in all lanes; under this scenario, the efficiency of each lane would remain constant since this is a relative measure of efficiency compared to the rest.

Using this framework, there is a new form of direct feedback for the competitiveness of rates that is not biased by changes in the market which should affect all lanes equally.

In practice, this framework for efficiency rating can be expanded to include more attributes for each lane, the specific date in which a measurement is taken and to compare the realizations of contracts given by several companies at a particular point in time. Moreover, under this framework it is also possible to get ratings of efficiency of not only lanes, but also specific instances spot rates obtained by companies using the lanes.

## 6.2 ANALYSIS OF SYSTEM-WIDE EFFICIENCY

In the same way in which we analyzed the behavior of transportation costs in Section 5, we can measure the behavior of efficiency of whole the system through time. If we graph the weighted average of the DEA efficiency values for the same clusters obtained in section 4.4 using the same framework and weights used in section 5.2 we obtain an estimate of the average relative efficiency of the whole system. Likewise, another approach was taken, consisting of taking the Gini coefficient on the relative efficiency of all lanes for each month; the result is a measure of statistical dispersion for the system wide efficiencies through time. These two calculations (Figure 6-3) yield separate indexes to that created in section 5, which convey complementary information about efficiency.

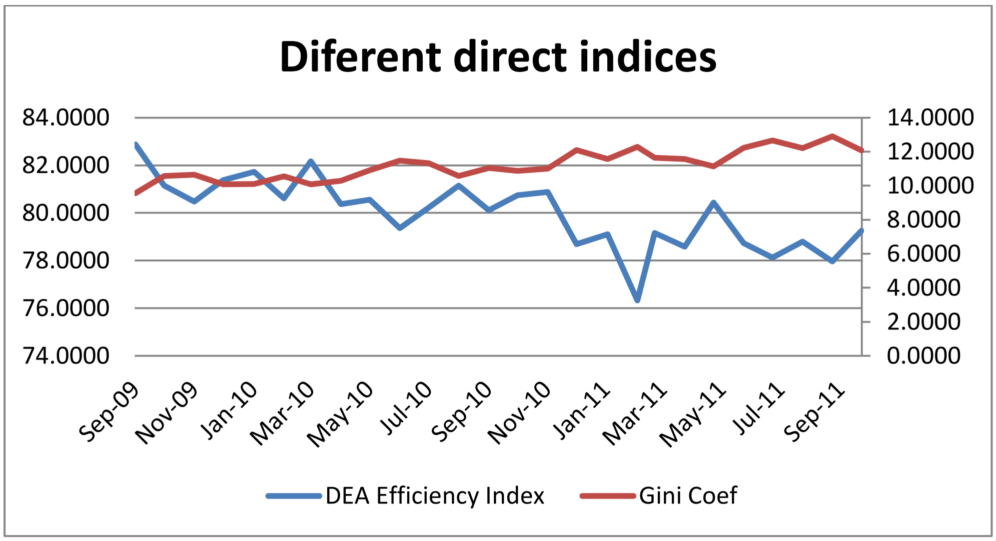


**Figure 6-3– Other efficiency Indexes**

The interpretation of Figure 6-3 would be that of overall system efficiency going down, while the dispersion of efficiencies is going up. In other words, the inequality in rate efficiency is increasing, while average efficiencies are also decreasing. Such an indicator can be used as an assessment of the degree of competition in the overall system.

Finally, another measure of interest would be that of how the relative efficiencies of lanes are geographically distributed. To show this, lanes are separated into 5 groups, depending on their efficiency, with group 1 being the least efficient and 5 being the most efficient. These groups can be seen in Figure 6-5 through Figure 6-8. Finally, Figure 6-9 shows the lanes that showed the most variability through time.

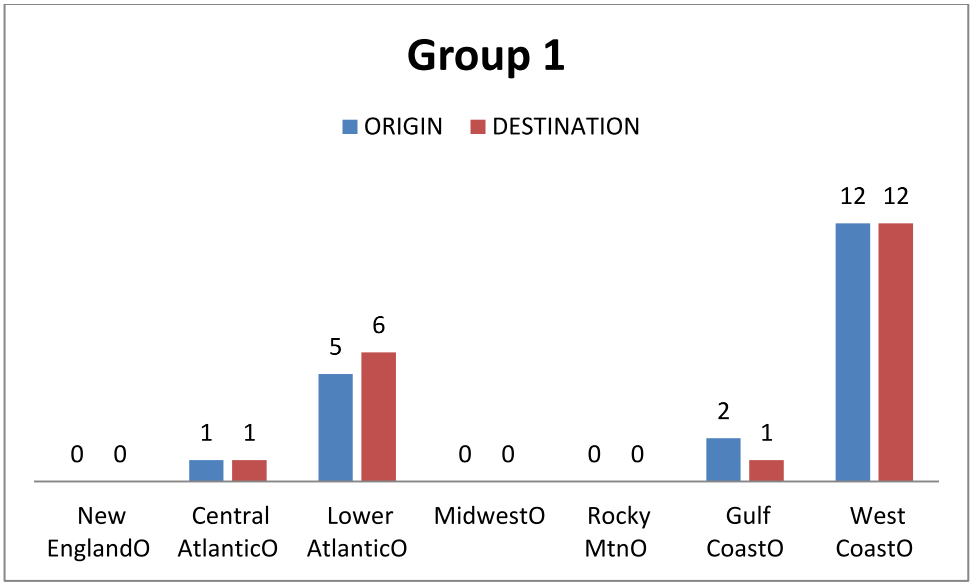


**Figure 6-4 – DEA relative efficiency (Most inefficient lanes)**

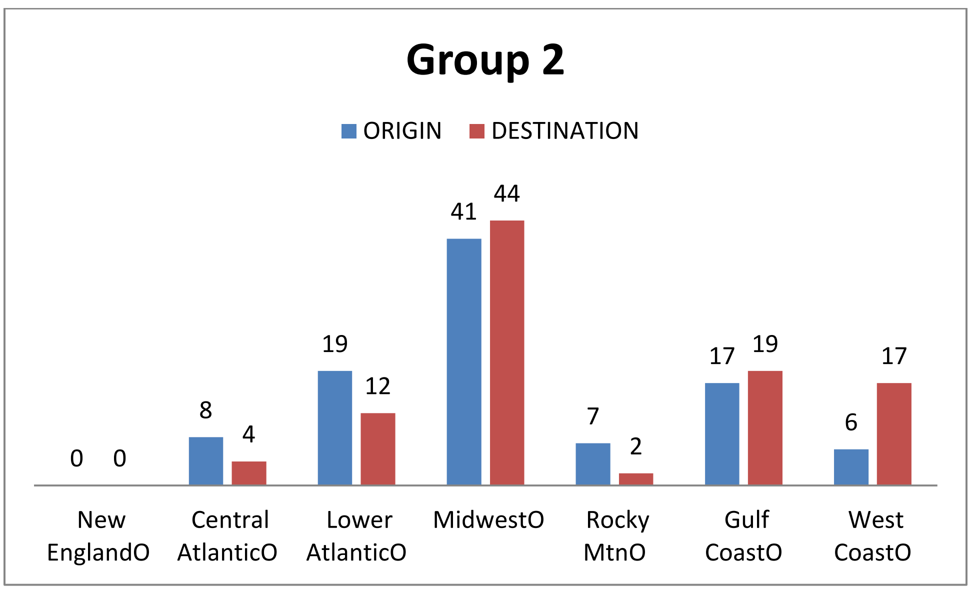


**Figure 6-5 – DEA relative efficiency (Moderately inefficient lanes)**

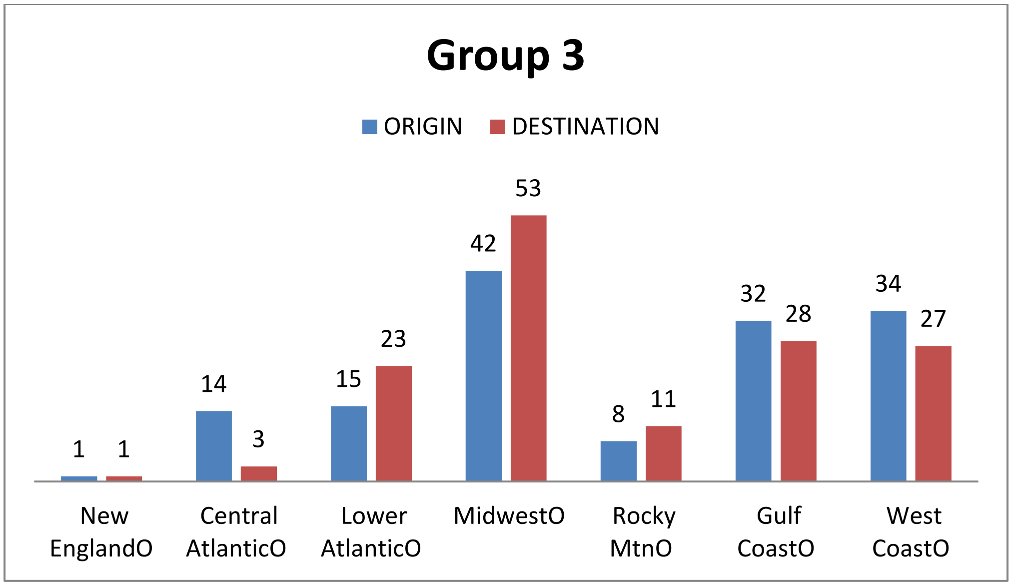


**Figure 6-6 – DEA relative efficiency (standard efficiency lanes)**

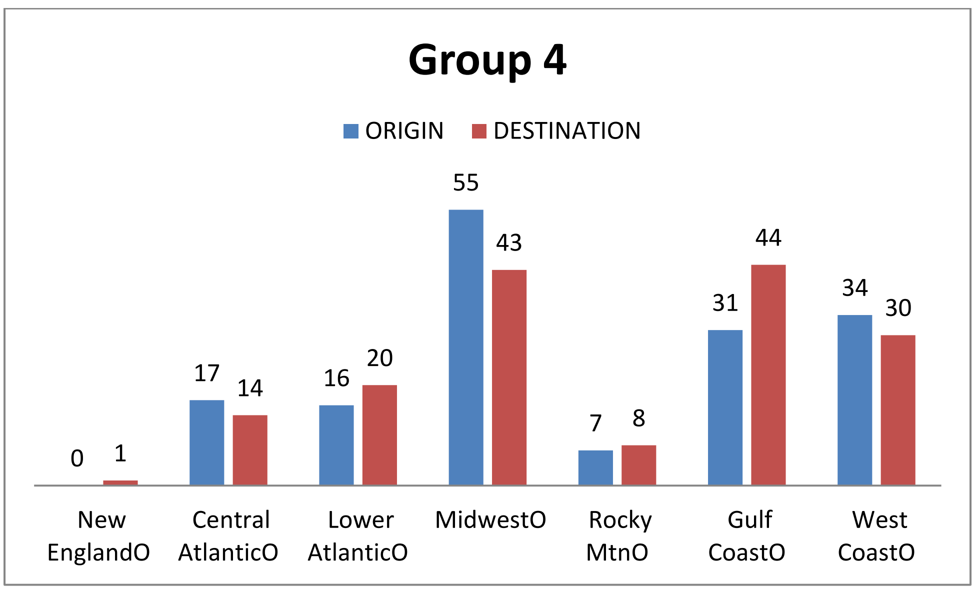


**Figure 6-7 – DEA relative efficiency (Moderately efficient lanes)**

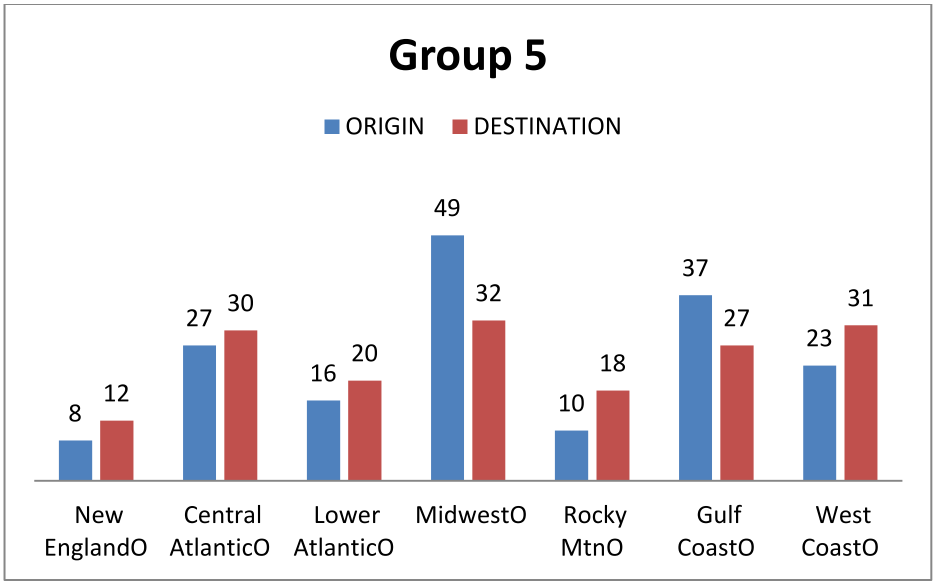


**Figure 6-8 – DEA relative efficiency (Most efficient lanes)**

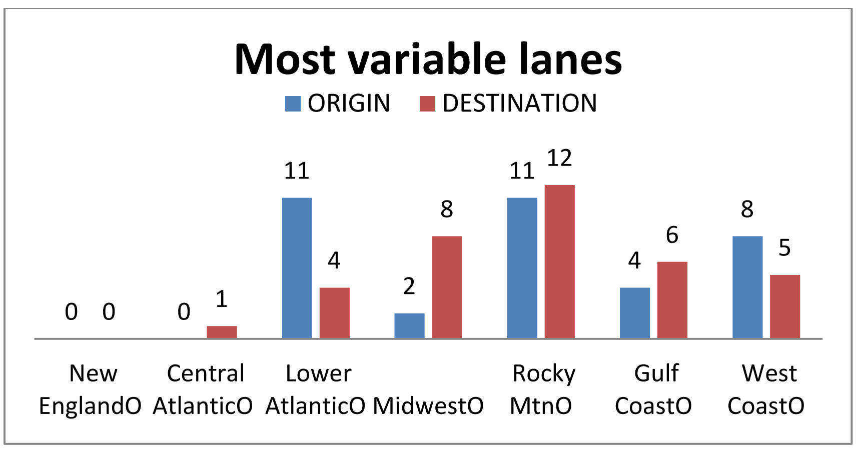


**Figure 6-9 – DEA relative efficiency (Most variable lanes)**

The previous groups are segmented through the use of a hierarchical clustering algorithm (Wards lineage). The results show that, when aggregated at a macro-level, such efficiency ranking effort may convey new information regarding the general distribution of the most efficient or inefficient lanes.

# 7 FUTURE WORK AND APPLICATIONS

The results obtained in the previous section have a considerable potential to create significant value through obtaining information at both the macro/system-wide level and the individual case for each lane. The principles and methodology for data aggregation and reduction can be applied to specific cases depending on the volumes of data. Ultimately, the data obtained can be used for various applications. This section details the development of the data structure and volumes of data that would be desirable for future implementations. A methodology for data collection is then illustrated through partnership with logistics companies, carriers and shippers. Finally, specific ways in which the information and results obtained throughout the current reports methodology and its expansion are detailed.

## 7.1 DEVELOPMENT OF A SHARING PLATFORM

In order to capture and process the relevant information for future analysis, research and implementations, an appropriate data structure must be defined. The right structure will facilitate data collection as well as aggregation of the information at several geographical levels of detail. In this project lanes are defined and aggregated using the main US statistical areas as a point of reference for origin/destination; in reality shipments can occur from more specific locations making the use of three-digit-zip-codes (3DZC) or zip-codes to identify the origin/destinations more desirable. The structure used for handling data throughout this project is shown in Figure 7-1; this can be used as a base for expanding into a more large scale implementation. A more detailed description of the tables and specific fields of the database can be found in Appendix F.

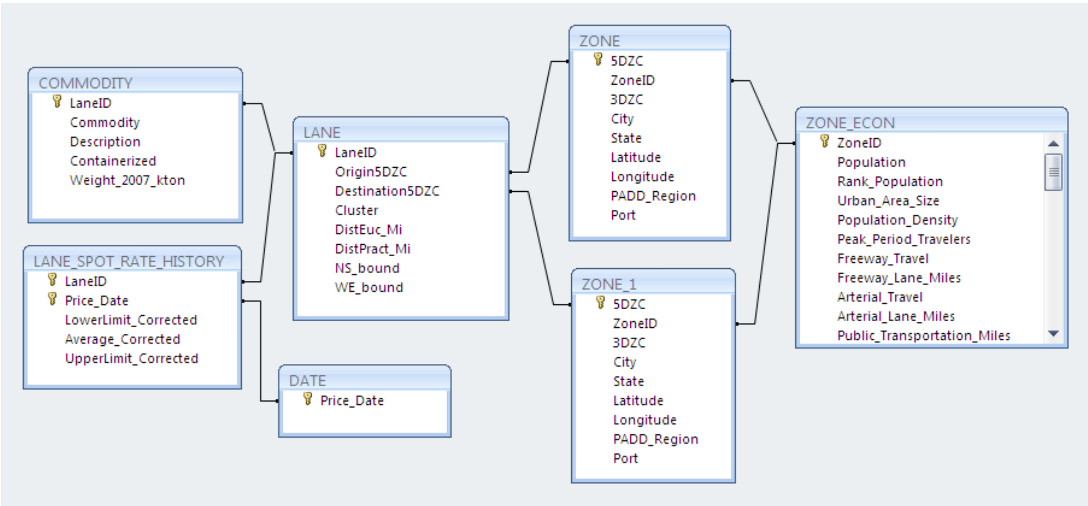


**Figure 7-1– Database structure**

As seen in Figure 7-1, each lane is associated to specific attributes that fall into distinct categories. For instance, each lane has a given quantity of goods of different kinds transported through it; likewise, lanes are associated with distinct regions, statistical areas and 3DZC, each of these with a given set of attributes; lanes are also assigned specific attributes like general heading and distance; finally, each lane will have a spot price for any given date in which these are recorded. The last part of the database, rate history, will be continuously updated, whereas the rest of the database will be relatively static. From this database, analysis can easily be made, with data aggregation at any level that is desired. The capability to aggregate data at various levels is an important feature if the information is used for research and for calculating indexes and specific rating metrics for each lane.

## 7.2 PARTNER WITH LOGISTICS COMPANIES AND CASE STUDY

To obtain the maximum benefit from future expansions in this project, large volumes of data are critical. A historical record of lane rates that spans several years would be a primary starting point. Nonetheless, to attain greater benefits, real time information on spot and contract rates must be available. This cannot be achieved without the collaboration of a large set of contributing companies, from carriers, shippers and 3PL companies.

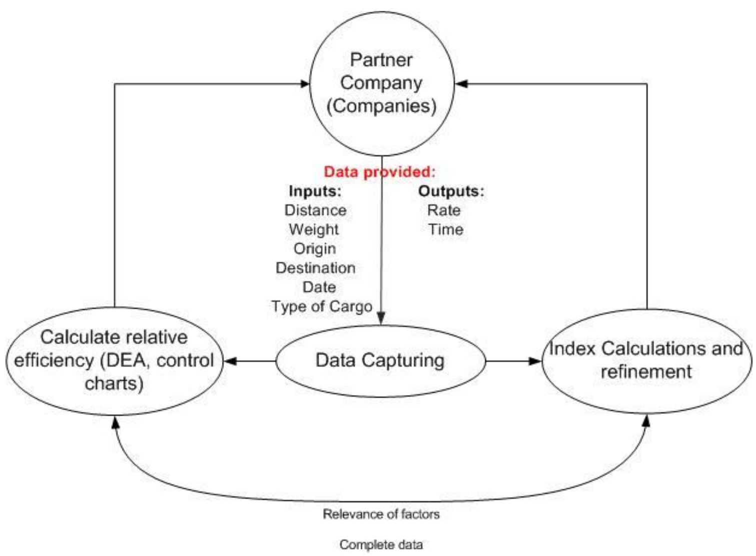


**Figure 7-2 – Information exchange and collaboration framework**

A framework is proposed in Figure 7-2through which companies would share their information and receive immediate feedback on the efficiency of specific lanes. Likewise, the information would be immediately anonymized, stored and become available for future benchmarking and calculation of global and local indexes.

## 7.3 FUTURE POTENTIAL AND APPLICATIONS

The application of a global index for estimating the status and direction of logistics efficiency in a macro level can have significant implications to the way in which business is made. If one or several indices are developed, then the direction of the market can be predicted and appropriate measures can be taken for contract writing and hedging. Likewise, if one specific lane is shown to be significantly less efficient than similar lanes, then the reasons for the discrepancy and the appropriate correction measures can be used. This can be used by logistics companies verifying the competitiveness of the rates being contracted. Some specific applications are further detailed in the following sections.

### 7.3.1 DIRECT BENCHMARKING THROUGH TIME

One common problem that logistics companies confront is that of estimating the real efficiency of their logistics operations outside of their companies or supply chains. Specifically, external points of comparison are limited and the level of detail of the information is very poor. While companies can access information for the behavior of rates in the specific lanes, this information is usually aggregated through a time period and inadequate for detailed comparisons.

Likewise, the most common comparison for the efficiency of a lane is that of comparing it against itself and to the current market average. However, there's also the possibility of benchmarking a lane against the behavior of lanes with a high degree of similarity to itself (such as those with geographically close origins and destinations), or to relate the lane to the general trend observed in the market. Such comparisons can be made by a direct contrast between the rates observed several lanes or by a relative measure of efficiency such as the one obtained through DEA.

### 7.3.2 VARIOUS PERFORMANCE MEASURES

When rating the efficiency of a lane, we note that rates provide a useful starting point which is globally utilized and of general interest to most companies. However, a lot of interest has been placed of alternative measures of efficiency such as time of delivery, variability and more recently environmental impact. Note that if further information (in addition to rates) is obtained, then further efficiency ratings can be calculated.

For instance, consider the case of environmental performance obtained in a particular lane or by a particular carrier. Clearly, there exist other factors that will also affect the performance of environmental impact measured in a lane-by-lane basis. Some of the factors affecting this can be:

- Age of the fleets
- Weather conditions

- Traffic in route
- Traffic at O/D
- Speed of travel
- Percentage of capacity utilized

An advantage of the framework developed throughout this study is that it can be generalized to deal with any performance measure of interest. The usage of DEA is particularly attractive in this case since it gives a quick and unbiased ranking at any level of detail that it is used.

### 7.3.3 CREATION OF A FUTURES MARKET FOR LOGISTICS AND TRANSPORTATION

Perhaps the most attractive application of an index for logistics and transportation is that of facilitating the creation of a market for futures and other logistics related derivatives. Futures contracts have been extensively used in other areas such as agriculture for hedging and reducing exposure to risk. It is widely known that the use of futures can be used for hedging even when the securities being traded have some correlation to the asset requiring a hedge such as the case of a minimum variance hedge (Luenberger, 1998). A logistics index and a division of lanes into similar clusters can be used as a starting point for identifying lanes that are appropriate for such hedging. As seen in section 5.1, clusters can be created through the use of statistical methods which are representative of the characteristics of a subset of lanes. Moreover, these clusters have been shown to have a certain historical correlation with each other. This information can be used for the pricing, purchase and writing of futures contracts.

To illustrate the use of logistics futures contracts for hedging, take the following example. Suppose that you are looking to make a shipment from the west coast to the Midwest. Unfortunately, there exist no futures contracts for the lane you are considering to use; moreover, the shipments made to this location are highly variable and the risk of a sudden increase in prices could make a significant impact on profits. Using the clusters and information developed in Section 5, we notice that the lane we are attempting to hedge belongs to cluster 3 and its behavior is consistent to that of the cluster; therefore, futures contracts within this cluster could be used to create a near perfect hedge.

Now suppose that no futures contracts for lanes within cluster 3 exist; even under this situation, a hedge can still be created. We notice that cluster 3 is highly correlated to cluster 0 as seen in **Error! Reference source not found.** and after some thought; we chose to purchase future ontracts from lanes in cluster 0 to create the best possible hedge. This action would effectively reduce the standard deviation of the costs observed in the future from 96 to 58; a 39% reduction.

**Table 7.1– Correlation between clusters**

| | *cluster0* | *cluster1* | *cluster2* | *cluster3* | *cluster4* | *Stdev* |
|---|---|---|---|---|---|---|
| cluster0 | 1 | | | | | 57.4 |
| cluster1 | 0.5595 | 1 | | | | 81.1 |
| cluster2 | 0.7373 | 0.9456 | 1 | | | 66.5 |
| cluster3 | 0.8222 | 0.2070 | 0.4537 | 1 | | 96.1 |
| cluster4 | 0.6850 | 0.9190 | 0.9294 | 0.3256 | 1 | 118.7 |

Ideally, logistics costs will be hedged by using futures contracts in other lanes. Nonetheless, even in the absence of a direct futures market for logistics resources, the clusters obtained for transportation lanes can be used as a reference for finding good hedging alternatives by comparing the behavior of the cluster to other economic indicators as shown in section 5.3.

### 7.3.4 AWARENESS OF SYSTEM STATUS AND VISIBILITY

One final application for the framework being proposed is that of providing further visibility to all companies involved in logistics and transportation. However, this visibility does not need to be limited to only rates and their behavior; another important outcome can be obtained from a large scale aggregation of data. That is: An estimation of the status for the entire logistics network.

As it was mentioned in previous sections, the trade surplus/deficit of a given zone plays a role in determining costs. Unfortunately, this cannot be estimated from publicly available data to the required level of detail nor on a month by month basis. However, if sufficient data is available, this can be estimated indirectly through the amount of outstanding contracts and the incoming/outgoing trucks at any given location for a given date. The availability of this information will allow logistics companies to more effectively manage their resources, plan their routes and ultimately reduce the amount of empty-miles traveled. Furthermore, if an approximate number of trucks currently stationed in each city are known, effective measures to keep the availability of trucks throughout the US at efficient levels can also be taken. The potential for further research on how this data can be used is large.

This outcome can greatly enhance the capacity of logistics companies to use fewer resources for moving their cargo, making them more profitable, as well as helping them achieve better rankings in terms of their resource usage and environmental impact. Better visibility of the entire network is probably the one of the greatest aids for achieving a green logistics network.

### 7.3.5 GREEN INDEX

In the past years, several environmental issues in logistics have been getting more relevant such as CO2 emissions, air pollution, road congestion and energy conservation. These problems need to be addressed by companies involved in logistics if they want to succeed in their business.

Companies have started to develop indexes which help them measure how “green” their operations are but there is not a general consensus about how to do it yet. “Green Logistics” is defined by (Rodrigue, Slack, & Comtois, 2012) as:

> *Supply chain management practices and strategies that reduce the environmental and energy footprint of freight distribution. It focuses on material handling, waste management, packaging and transport.*

The majority of the current indexes try to measure the efficiency of “forward” logistics but in green logistics the need to measure “reverse logistics” becomes even more important. Reverse logistics relates to the transportation of used material and waste from the consumer to the disposition destinations, as opposed to the traditional producer-consumer flow of goods. The index that we are proposing can include a way to measure how efficient a company is regarding its reverse logistics process through DEA as detailed in section 6.1.3. Likewise, it can aid companies to better plan their operations as to minimize the impact on the environment through better system visibility.

For example, a combined index could be implemented which includes green indicators that help measure how eco-efficient a supply chain is. The green index could include not only the traditional logistics activities such as distribution, networks redesign, shortening travel distance and shipment consolidation; but also the reduction of diesel consumption per kilometer or ton moved, C02 emissions reduction, etc. The information gathered could include the percentage of alternative energy such as “biodiesel” that is being used, the percentage of water used to wash the fleet and percentage of water recycled, tires and oil disposition could also be included as standard indicators to be used as benchmarking.

Finally, the green index could also include the percentage of goods moved by truck, rail or ship. It is well known that ships are about 10 times as fuel-efficient as trucks and as twice efficient as train. This new index which includes environmental information could be used by current programs such as Smartway Transport Partnership (EPA, 2012). This is a program in which freight carriers and shippers commit to benchmark operations, track fuel consumption and improve performance annually. This index would help assess freight operations to reduce the footprint and become more fuel-efficient.

# 8 CONCLUSIONS

Through this research, it was seen that the definition of logistics efficiency at the supply chain level is ambiguous; moreover, there are no tools available for monitoring the efficiency of logistics resources outside of a specific company or supply chain. Therefore, research was conducted to address the issue of how to monitor efficiency of logistics resources as well as estimating the overall status and potential trends in logistics costs.

Throughout this report a methodology for analyzing large scale information on logistics efficiency was developed. It was shown that statistical methods such as regression analysis and data mining can be used to obtain information on the relevant factors affecting logistics efficiency and therefore conduct a targeted approach to understand their behavior and relationship to one another. More importantly, it was shown that the behavior of the entire system can be understood through the use of the appropriate techniques. The information obtained through statistical analysis can help to further segment the lanes under consideration and perform a targeted assessment of efficiency depending on the characteristics of the lanes.

The utilization of statistical techniques allows us to build upon the relevant factors to generalize our knowledge about the system. Taken further, this knowledge was used to develop a single performance indicator for the system, which has a strong potential for predicting trends in costs. Likewise, building upon the factors identified from statistical analysis it was possible to measure efficiency of specific instances by direct comparisons of similar lanes and relative efficiency ranking. We propose further applications for the use of these indices and metrics, expanding them towards: benchmarking of various logistics performance estimates; contract writing, negotiation and the creation of a futures market for logistics contracts; and to provide visibility of the full logistics network through data aggregation and estimation.

Expansions and learning outcomes:

The current report proposes a framework for analyzing and utilizing the information available to achieve better management of logistics resources. Although the current project was limited by the availability of data, the framework developed can still be applied to much larger datasets, expecting better results. We propose that the methodology used can be expanded in future projects to obtain further information about the global logistics system dynamics.

As a potential expansion to the model it is desirable to consider much larger amounts of historical data for the creation and validation of an index for transportation and logistics costs. In this project, good results were obtained using historical data spanning 26 months. Nonetheless, using more historical data, a better estimate for the correlation of our index and other economic indicators can be achieved.

Under larger data availability, the clustering approach to create representative groups can also be modified. In this project, it was assumed that the characteristics of lanes relevant for predicting cost would be good alternatives for clustering. In subsequent investigation it would be beneficial to investigate the creation of different clusters by considering the behavior of costs for the various lanes through time. Likewise, when analyzing the factors relevant for predicting logistics costs through regression analysis two changes would be worthwhile to consider: One is the further segmentation of the PADD regions of the US; the other one is the analysis of the behavior of logistics costs as a time series problem, identifying seasonality factors.

Ultimately, the more detailed analysis of the data will lead to results of greater confidence and implementation of the various applications to the use of logistics efficiency monitoring.

# References

Ahumada, O. (2003, December). Efficiency Measure for Inbound Logistics Performance (Master Thesis). Arizona State University, Tempe, AZ.

Bureau of Transportation Statistics. (2010). BTS | Transportation Services Index. Retrieved September 6, 2010, from http://www.bts.gov/xml/tsi/src/index.xml

Cass Information Systems, Inc. (2010). Cass Information Systems, Inc. - Freight Index. Retrieved September 29, 2010, from http://www.cassinfo.com/frtindex.html

Charnes, A., Cooper, W. ., Golany, B., Seiford, L., & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. Journal of Econometrics, 30(1–2), 91–107. doi:10.1016/0304-4076(85)90133-2

Davis, P., & Drum, F. (1996). Presented at the Council of Logistics Management Annual Conference, Oak Book Factor.

Hausman, W. H., Lee, H., & Subramanian, U. (2005). Global Logistics Indicators, Supply Chain Metrics, and Bilateral Trade Patterns. SSRN eLibrary. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=869999

Huan, S. H., Sheoran, S. K., & Wang, G. (2004). A review and analysis of supply chain operations reference (SCOR) model. Supply Chain Management: An International Journal, 9(1), 23–29. doi:10.1108/13598540410517557

Institute for Supply Management. (2010). ISM - ISM Reports on Business. Retrieved September 29, 2010, from http://www.ism.ws/ISMReport/?navItemNumber=4892

Keebler, J. S., Manrodt, K. B., Durtsche, D. A., & Ledyard, D. M. (1999). Keeping Score: Measuring the Business Value of Logistics in the Supply Chain. Council of Logistics Management.

Luenberger, D. G. (1998). Investment science. Oxford University Press.

Molinero, C. M., & Woracker, D. (1996). Data Envelopment Analysis: OR Insight, 9(4), 22–28. doi:10.1057/ori.1996.21

Rodrigue, J.-P., Slack, B., & Comtois, C. (2012). Green Logistics. THE GEOGRAPHY OF TRANSPORT SYSTEMS. Retrieved from http://people.hofstra.edu/geotrans/eng/ch8en/appl8en/ch8a4en.html

Ryan Barnes. (2010). Economic Indicators: Purchasing Managers Index (PMI). Retrieved August 30, 2010, from http://www.investopedia.com/university/releases/napm.asp

Shanghai Shipping Exchange. (2009, October 16). Shanghai Containerized Freight Index. Retrieved August 26, 2010, from http://en.chineseshipping.com.cn/html/scfi.html

SmartWay | US Environmental Protection Agency. (n.d.). Retrieved April 4, 2012, from http://www.epa.gov/smartway/index.htm

The Baltic Exchange. (2010). The Baltic Exchange - Home Page. Retrieved September 29, 2010, from http://www.balticexchange.com/

The World Bank. (2010). Trade Logistics & Faciliation - Logistics Performance Index (LPI). Retrieved August 6, 2010, from http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTTRANSPORT/EXTTLF/0,,contentMDK:21514122~menuPK:3875957~pagePK:210058~piPK:210062~theSitePK:515434,00.html

TransCore. (2010). TransCore. Retrieved from http://www.transcore.com/

U.S. Census Bureau. (2008, March). Metropolitan and Micropolitan Statistical Areas of the United States and Puerto Rico November 2008. U.S. Census Bureau. Retrieved from http://www.census.gov/geo/www/maps/msa_maps2008/msa2008_previews_html/cbsa_us_wall_1108.html

U.S. Energy Information Administration. (2011). States in Each Region. U.S. Energy Information Administration, U.S Department of Energy. Retrieved from http://www.eia.doe.gov/oil_gas/petroleum/data_publications/wrgp/padd_page.html

U.S. Energy Information Administration. (2012, April 2).Gasoline and Diesel Fuel Update. Retrieved October 10, 2012, from http://www.eia.gov/petroleum/gasdiesel/

UCLA Anderson School of Management. (2010). UCLA Pulse of Commerce Index (PCI) Definition. Retrieved April 20, 2011, from http://www.ceridianindex.com/about-the-index/

# Appendix A – Literature Review

# Literature Review on Logistics Metrics

## 1. Introduction

One of the main premises of this project is to identify and assess the potential factors that influence the logistic performance of current U.S. land transportation networks. Through this study, a methodology will be developed to measure and benchmark the logistics performance of these networks. This measurement will be based on specific lanes, the main commodities transported through them and the different factors affecting their performance (both external and internal to the transportation networks). The first part of the study will consist of identifying the main land transportation lanes, as well as the commodities that they transport. The second part consists of applying statistical analysis methods to the collected data in order to determine the factors that are most relevant to the performance of the specific lanes. It is not yet clear which factors are the most influential; potential factors may include rate per ton per mile, utilization of capacity, delivery accuracy and others. The final list of factors may change as the objective of the project is defined. Statistical methods will be applied to find significant factors among a candidate factor list.

In this document a literature review is conducted to summarize the existing studies related to common logistics metrics. The studies summarized in this review will provide an overview of the current status of researches related to logistics performance. Furthermore, a list of existing performance indices is explained, which will provide a sense of how companies in industry gauge their logistics operations. This document is organized as follows: A review on existing studies is presented first followed by a general overview of commonly used indices; finally a tentative conclusion is drawn.

## 2. Existing studies

This section reviews studies related to logistics performance. A measurement of logistics performance is important and critical to apply strategies to the logistics system. Several researchers have attempted to demonstrate the importance of logistics performance measurement (Tracey 1998; Fawcett & Cooper 1998; Keebler et al. 1999). The performance of logistics can influence a manufacturer's ability to satisfy its customers and its overall operations. Thus this topic attracted great attention from the researchers. However, the definition of logistics performance is not yet clear. Besides, the context of logistics performance also changed significantly as the business environment changed rapidly during last few decades.

The definition of logistics performance is ambiguous, which is because when one talks about logistics performance, the definition of the word varies significantly depending on the scope in which it is treated. For example, the meaning of "logistics performance" from a company's perspective is different from that from a region or a country's perspective. Nevertheless, most studies focus on the logistics performance of a company or an individual supply chain as it may be easier to collect information for a company than that of a region or country. Thus, most definitions about "logistics performance" have a corresponding limit of vision. Due to this reason, we also investigate further to see what factors are considered to formulate a logistics performance metric.

Chow, Heaver, and Henriksson (1994) conducted a review on the definition of logistics performance. Measures are categorized to "Hard" and "Soft", in which the former includes income or accounting figures and the latter includes measures such as customer satisfaction ratings. It is also common to call these two category measures as financial factors and non-financial factors, respectively. Caplice and Sheffi (1994) divided logistics measurement into three different dimensions: utilization, productivity and effectiveness. They also compared different logistics metrics with eight different criteria: validity, robustness, usefulness, integration, economy, compatibility, level of detail and behavioral soundness. This study provides a comprehensive framework to understand a logistics performance. A more recent review on this topic was authored by Gunasekaran & Kobu (2007). A question asked by this study was: as the business environment change so rapidly during last decade, are the logistics performance metrics developed previously still capable to capture all the characteristics of a modern logistics system? They indicated that some traditional Performance Measurements (PM) such as price, time and cost may be questioned for their validity. However, they cannot be eliminated due to some historical reasons.

There are some factors commonly considered by different researchers. Table 1 (Table 3 in Angappa Gunasekaran and Kobu 2007) shows a summary of metrics used in existing studies, where the literatures were also divided into different groups according to their perspectives. From the table below we can see that the meaning of "logistics" is not restricted to transportation in the literature. Factors like warehousing, inventory costs are also considered. However, in our study, transportation is the main concern. The inventory cost seems to draw the greatest interest from the researchers, which is indirectly related to transportation accuracy, flexibility and costs.

Ahumada (2003) evaluated several logistics performance metrics for a company level transportation system. Several existing metrics as well as a self-formulated metric were compared in that study and a framework for evaluation was built. The framework from this study is potentially helpful to our study. A study that evaluates the logistics performance of a country (Hausman et al. 2005) also caught our attention. This study was supported by World Bank, and the resulting index is updated yearly. Compared to other studies' focusing on company level, this study provides insights on how to build logistics performance measure for a region.

As we can see from the literature, studies with scope beyond a company or a supply chain level are very limited. To the best of our knowledge, there is neither a study dedicated on the logistics performance of a particular lane, nor a methodology that can be used to evaluate the logistics performance of a lane.

**Table 1 Metrics used to measure performance in SCM systems and their relations to categories and factors suggested by researchers (**Table 3 in Angappa Gunasekaran and Kobu 2007**)**

| | A | | | | B | | | | C | | | | | D | | | E | | F | | G | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metrics | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 1 | 2 | Total | Percentage |
| 01 Accuracy of scheduling | | X | | | X | | | | X | | | | | | | X | | X | X | | X | | 7 | 32 |
| 05 Bid management cycle time | | X | | | X | | | | X | | | | | X | | | | X | X | | X | | 7 | 32 |
| 06 Capacity utilization | | X | X | | | X | | | | | X | | | | X | X | | X | X | | X | | 9 | 41 |
| 07 Compliance to regulations | | X | X | | | | X | | X | X | | | | X | | | | X | X | X | | X | 10 | 45 |
| 08 Conformance to specifications | | X | | X | | | X | | | X | X | | | | | X | | X | X | | | X | 9 | 41 |
| 18 Delivery reliability | | X | | X | X | | X | | | | | | X | | | X | | X | X | | | X | 9 | 41 |
| 24 Forecasting accuracy | | | | X | X | | X | | X | | | | | | X | X | | X | X | | | X | 9 | 41 |
| 29 Inventory costs | X | X | | | | X | | X | X | X | X | X | X | | | X | X | | X | | X | X | 14 | 63 |
| 33 Labor efficiency | | | X | | | X | | | | | X | | | | X | X | | X | X | | X | | 8 | 36 |
| 35 Lead time for procurement | | | | X | X | | | | | X | | | | | X | | | X | X | | X | | 7 | 32 |
| 36 Lead time manufacturing | | X | | X | X | | | | | | X | | | | X | X | | X | X | | X | | 9 | 41 |
| 39 Obsolescence cost | X | | | X | X | | X | | X | | | | | | | | X | | X | | | X | 8 | 32 |
| 44 Overhead cost | X | X | | | X | X | | | X | | X | | | | | X | X | | X | | | X | 10 | 42 |
| 46 Perceived quality | | | | X | | | X | X | | | | | X | | | | | X | | X | | X | 7 | 32 |
| 47 Perceived value of product | | | | X | | | X | | | | | | X | | | | | X | | X | | X | 6 | 27 |
| 50 Process cycle time | | X | | | X | | | X | X | | X | X | | | X | X | | X | X | | X | | 11 | 50 |
| 51 Product development time | | X | X | | X | | | | X | | | | | X | | | | X | X | | X | | 8 | 36 |
| 54 Product/service variety | X | | X | X | | | X | X | X | | | | | X | | | | X | | X | | X | 10 | 45 |
| 55 Production flexibility | | X | X | | X | X | X | X | | | X | | | | | X | | X | | X | X | | 11 | 50 |
| 62 Return on investment | X | | | | | X | | | X | | | | | X | | | X | | X | | | X | 7 | 32 |
| 63 Selling price | X | | | X | | X | X | | | | | | X | X | | | X | | X | | | X | 9 | 41 |
| 68 Stock out cost | X | | | X | | | X | X | | | | | X | | | | X | | X | | | X | 8 | 32 |
| 71 Supply chain response time | | X | X | | X | | | X | X | | | | | X | X | X | | X | X | | X | | 11 | 50 |
| 76 Transportation cost | X | | | | | X | | | | | | X | | | X | X | X | | X | | | X | 8 | 32 |
| 77 Value added | X | | | X | | X | | | | | X | | X | | X | X | X | | X | | | X | 10 | 45 |
| 81 Warranty cost | X | | | X | | | X | | X | | | | | | | | X | | X | | | X | 7 | 32 |
| Total | 10 | 13 | 7 | 13 | 12 | 9 | 12 | 7 | 13 | 4 | 9 | 3 | 7 | 7 | 9 | 14 | 9 | 17 | 22 | 5 | 11 | 16 | | |
| Percentage | 38 | 50 | 27 | 50 | 46 | 35 | 46 | 27 | 50 | 15 | 35 | 12 | 27 | 27 | 35 | 54 | 35 | 65 | 85 | 19 | 42 | 61 | | |

**A**. Balance score perspectives: 1. Financial, 2. Internal process, 3. Innovation and improvement, 4. Customers. **B**. Components of performance measures: 1. Time, 2. Resource utilization, 3. Output, 4. Flexibility. **C**. Location of measures in supply chain links: 1. Planning and product design, 2. Supplier, 3. Production, 4. Delivery, 5. Customer. **D**. Decision level: 1. Strategic, 2. Tactical, 3. Operational. **E**. Financial base: 1. Financial, 2. Nonfinancial. **F**. Measurement base: 1. Quantitative, 2. Nonquantitative. **G**. Traditional vs. Modern: 1. Function based, 2. Value based.

## 3. Existing Indices

Besides the studies related to logistics performance, we also look for indices that are commonly used in the industry. By reviewing these indices, we identify what are concerned in the industry. The methodology and data set used by these indices may potentially be useful to our study. These indices are categorized into three different groups. The first group contains indices related to maritime transportation, while the second group contains indices mostly related to terrestrial transportation. The third group of indices is not directly related to transportation, and hard to fit into previous mentioned categories.

The first group of indices is for marine transportation, we will go through them one by one.

1. **Baltic Dry Index** (The Baltic Exchange 2010)

   The Baltic Dry Index is a daily average of prices to ship raw materials. It represents the cost paid by an end customer to have a shipping company transport raw materials across seas on the Baltic Exchange, the global marketplace for brokering shipping contracts. It measures the demand to move raw materials and precursors to production. The index is quoted every working day at 13:00 London time. It takes 20 different routes throughout the world for various materials and averages them into one index

2. **Shanghai Containerized Freight Index (SCFI)** (Shanghai Shipping Exchange 2009)

   The SCFI is published by Shanghai Shipping Exchange. It shows the ocean freight and surcharges of individual shipping routes on the spot market. 15 individual shipping routes are included, which are worldwide, not only limited to those related to China. The index is updated weekly.

3. **Logistics Performance Index – LPI** (The World Bank 2010)

   This index is the result of the aforementioned study supported by World Bank. It evaluates the logistics efficiency of each country. Unlike the previous two indices, the LPI is not a single index but a group of indices give measurement in different dimensions. The LPI is annually updated, and is available from World Bank. The detailed evaluation methods can be found in the study from Hausman et al. (2005).

The second group includes inbound freight transportation, mostly terrestrial.

1. **CASS Freight Index** (Cass Information Systems, Inc 2010)

   This is a monthly Volume Index of Freight Expenditures and Shipments that is based upon transportation dollars and shipments of Cass Information System's clients. The company CASS processes over $17.5 billion in annual freight payables. It uses January 1990 as its base month and it is updated with monthly freight expenditures and shipment volumes. The purpose is this index is to compare levels of shipment activity on a month to month basis. This index does not equate to CASS' overall processing volumes but to the volume levels of CASS' customer base. The intelligence obtained from the data used to generate this index is limited. Potentially, more sophisticated analysis can conduct on this data to obtain more intelligence.

2. **Transportation Service Index (TSI)** (Bureau of Transportation Statistics 2010)

   This index is created by the U.S. Department of Transportation (DOT), Bureau of Transportation Statistics (BTS), measures the movement of freight and passengers. The index, which is seasonally adjusted, combines available data on freight traffic, as well as passenger travel that have been weighted to yield a monthly measure of transportation service output. The main transportation elements upon which the index is based on are the following: rail (passenger and freight), pipeline (petroleum and natural gas), transit, waterborne, trucking, and aviation (passenger and freight). One of the downsides of the BTS index is its lack of control over upstream data release, which in turn causes variation in the timely publication of the index. This is because BTS does not collect data by itself, but obtains them from upstream source. For example, the trucking data are obtained from American Trucking Association.

3. **Indices from Freight Transportation and Research (FTR) Associates** (FTR Associates 2010)

   Freight Transportation and Research Associates provides several indices related to the trucking industry, which includes freight trends, industry capacity, trucking costs, driver supply and other factors affecting their operations. FTR uses available data and proprietary calculations to forecast the upcoming demand for freight services. The items included in this report are the following:

   - **FTR Trucking Conditions Index** combines five trucking industry statistics into one metric measuring the overall health of the trucking industry.
   - **FTR Truck Loadings Index (Y/Y %)** calculates the percentage of the total population of trucks that is required to move the U.S. truck freight. In general, a figure above 90% indicates a tight market where the majority of the truck population is at work. A figure below 85% indicates a weak market where a significant portion of the truck population is idle.
   - **FTR Truck Utilization Rate** metric calculates the percentage of the total population of trucks that is required to move the U.S. truck freight. In general, a figure above 90% indicates a tight market where the majority of the truck population is at work. A figure below 85% indicates a weak market where a significant portion of the truck population is idle.

4. **Grain Transportation Report (GTR)** (USDA AMS 2010)

   The Agricultural Marketing Service (AMS) provided by United States Department of Agricultural (USDA) releases a weekly Grain Transportation Report and a quarterly summary. The report contains rate per mile per truckload, Truck availability, Truck usage and Future truck usage information. It reports these indices according to statistical regions, which are West, Rocky Mountain, South Central, North Central and East regions. The rate is reported in price, while all other indices are reported by a 1-5 scale. Besides these, fuel price and ocean rate to major export destinations (Bulk transportation, $/Ton) are also reported in their report.

5. **Manufacturers' Shipments, Inventories, & Orders** (U.S. Census Bureau 2010)

   The purpose of the Manufacturers' Shipments, Inventories, and Orders (M3) survey is to provide broad-based monthly statistical data on current economic conditions and indications of future production commitments in the manufacturing sector. The M3 is based upon data reported from manufacturing establishments with $500 million or more in annual shipments. The survey has been conducted monthly by the US Census Bureau since 1957. The report provides statistics on manufacturers' value of shipments, new orders (net of cancellations), end-of-month order backlog (unfilled orders), end-of-month total inventory (at current cost or market value), and inventories by stage of fabrication (materials and supplies, work-in-process, and finished goods).

The last group refers to miscellaneous indices that are not directly related to logistics.

1. **Purchasing Managers Index (PMI)** (Institute for Supply Management 2010)

   The PMI is an indicator of the economic health of the manufacturing sector. It is issued by the Institute for Supply Management (ISM), which is a non-profit group boasting more than 40,000 members engaged in the supply management and purchasing professions. The latest report can be found of ISM official website. An explanation of the calculation of PMI can be found from INVESTOPEDIA (Ryan Barnes 2010). The PMI is a composite index of five "sub-indicators", which are extracted through surveys to more than 400 purchasing managers from around the country, chosen for their geographic and industry diversification benefits. The five sub-indexes are given a weighting, as follows: Production level (.25), New orders (.30), Supplier deliveries (.15), Inventories (.10), Employment level (.20).

   A diffusion process is done to the survey answers, which come in only three options; managers can either respond with "better", "same", or "worse" to the questions about the industry as they see it. The resulting PMI figure (which can be from 0 to 100) is calculated by taking the percentage of respondents that reported better conditions than the previous month and adding to that total half of the percentage of respondents that reported no change in conditions.

2. **OPIS Transportation Fuel Index (TFI)** (OPISNET.COM 2010)

The Transportation Fuel Index delivers precise and comprehensive wholesale statistical data. Wholesale and retail data within the TFI is available within hours of a month's conclusion, and can be catalogued by region, state, metropolitan statistical area (MSA) or even by zip code. The OPIS TFI gives analysts an immediate picture of wholesale performance and trends. For gasoline, diesel, and jet fuel prices, this will end the frustration in working with out-of-date statistics. This index is an indirectly related to transportation. The most valuable part of this index is that it is prompt availability of historical data and forecast of near future.

3. **Delta Global Shipping Index** (Delta Global Advisors 2010)

   The Index is designed to measure the performance of companies listed on global developed market exchanges and currently consists of companies within the maritime shipping industry. This index tracks 30 companies from different business segments. Only companies that generate in excess of 80% of their revenues (significant) from the operating and/or leasing of seaborne ships which transport dry bulk, tanker, container, specialty chemical or LNG goods are included. Besides this, there are also restriction on the capital size of the companies and other aspects. Detail on these restrictions can be found from their company website. The Delta Global Shipping Index is re-constituted annually, with re-balancing occurring quarterly according to the rules of index construction. This index mainly measures the stock performance these companies and synthesize them into a single index. A similar index can be found is the Dow Jones Transportation Average (Dow Jones Indexes 2010).

The first two groups of indices listed in this subsection will most likely reflect the type of index developed through this study. The indices for marine transportation shares some similarity with this study, though have different subjects as ours. This is because the shipping routes included in these indices are point–to-point, while the "lane" under investigation in our study is also point to point. Also, the data used to generate these indices may be potentially useful for our study. Indices in the third group may not be so closely related, however, they provide some general guidelines of where to get data for this study. At the time we are writing the literature review, more and more indices are discovered. Although they are not included in this literature review, they can well fit into one of the three categories.

## 4. Additional Sources of Information: Combinatorial Auctions

There are various additional sources of information that could be used to develop a potential logistic performance metric. One of these is the use lane rates between shipper and carrier contracts. Specifically, one could use the resulting bids of combinatorial auctions (reverse auctions) to identify the specific average rates for specific carrier lanes; as well as to deduce other kind of information such as their relative importance, utilization, and availability. Once obtained, the data results from combinatorial auctions will provide us an additional insight into the analysis process of the different lanes.

The main objective of combinatorial auctions is to recognize the potential profits locked in transportation through the use reverse auctions. In this case, the carrier is the one who bids a transportation service price to the shipper based on different service packages formulation of their operations.  The shipper then considers the bid and selects the service packages that best fits *their* own operations (Sheffi 2004). This practice is known as combinatorial auctions, in which the different players bid and consider different service combinations to select the one that best fits their objectives.

There are several benefits from engaging in combinatorial auctions. Leading shippers take advantage of combinatorial auctions to save three to 15 percent of transportation costs while maintaining or increasing their service levels (Sheffi 2004). Also, the optimization framework inherent in determining the winners in these combinatorial auctions helps shippers achieve many corporate goals beyond minimizing transportation costs.  Lastly, the carriers benefit from offering service on lanes that will balance their networks.

One could possibly use the results of these auctions as a way to monitor important characteristics of each lane.  Combinatorial auctions might be a good application for the developed index in this project. The players involved in the reverse auctions could potentially benefit from a metric that would help them gauge their decisions on future bids. This might provide leverage when asking for data related to bids and auction results. Therefore, possible benefits packages might be needed in order to strike a deal with entities holding this kind of information.

# 5. Conclusion

The main objective of this literature review is to present a current outlook on the subject of logistics performance metrics and explore possible advancing direction of this study. The literature review focuses on two parts: one for theoretical studies and the other for practically applied indices. As part of the assessment, different papers were reviewed in order to define the current scope of logistics performance and its application.

By reviewing existing studies, it was found that the definition of the “Logistics Performance” is ambiguous and dependent on the context on which it is applied. The word in the context of a country or region has totally different meaning of that in the context of a single company. As the result, the research methods applied also differ from each other. On the other hand, we found that there is a strong demand on such kind of logistics performance measurement. This is because logistics performance is significant to the operation of the company to satisfy its customers and improve its overall performance.

The current project aims to develop a performance metric that can be used and applied by the general industry within their logistics operations. More specifically, the current project aims to evaluate the logistics performance in the unit of a lane. Thus, the result of this project will not be limited to any particular company, but can be adopted by all the participants along the lane.

For this reason, another aspect of the review was to present a glimpse of the indices/metrics currently used by companies to make logistics decisions. These indices were divided among three groups based on their basic application: marine, terrestrial, and miscellaneous. These indices give a general idea of the possible application to our own index/metric, and most importantly present possible sources of information that could be tapped for our own purposes. This literature review listed some of the typical ones, while there are a lot more similar indices in the market. However, most of them can find a spot in one of our three categorizations.

During the review process it was found that there is an opportunity to improve on the current index list to include other areas of performance measure. Specifically, there is an opportunity to develop a metric based on the transportation lanes that carriers and shippers use to operate their business. This indicator of the main transportation lanes could be used to assess the general trend of a particular lane's performance based on a variety of factors, which might include the following: reliability and utilization of the lane, availability of carriers, road conditions, etc. This index could be expanded to include time-based factors, such as seasonality, as part of its inputs. Nevertheless, further assessment of the practicality and relevance of this possible research extension needs to be defined before continuing the study.

## 6. Bibliography

Ahumada, O., 2003. *Efficiency Measure for Inbound Logistics Performance*. Master Thesis. Tempe, AZ: Arizona State University.

Bureau of Transportation Statistics, 2010. BTS | Transportation Services Index. Available at: http://www.bts.gov/xml/tsi/src/index.xml [Accessed September 6, 2010].

Caplice, C. & Sheffi, Y., 1994. A Review and Evaluation of Logistics Metrics. *The International Journal of Logistics Management*, 5(2), 11-28.

Cass Information Systems, Inc, 2010. Cass Information Systems, Inc. - Freight Index. Available at: http://www.cassinfo.com/frtindex.html [Accessed September 29, 2010].

Chow, G., Heaver, T.D. & Henriksson, L.E., 1994. Logistics Performance: Definition and Measurement. *International Journal of Physical Distribution & Logistics Management*, 24(1), 17-28.

Delta Global Advisors, 2010. Delta Global Shipping Index – Maritime Shipping Index. Available at: http://dgshippingindex.com/ [Accessed September 29, 2010].

Dow Jones Indexes, 2010. Dow Jones Transportation Average. Available at: http://www.djaverages.com/?view=transportation&page=overview [Accessed September 30, 2010].

Fawcett, S.E. & Cooper, M.B., 1998. Logistics Performance Measurement and Customer Success. *Industrial Marketing Management*, 27(4), 341-357.

FTR Associates, 2010. FTR Freight Focus: Trucking Update. Available at: http://www.ftrassociates.com/public/home/document.php?dA=dashboard [Accessed September 29, 2010].

Gunasekaran, A. & Kobu, B., 2007. Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995-2004) for research and applications. *International Journal of Production Research*, 45(12), 2819-2840.

Hausman, W.H., Lee, H. & Subramanian, U., 2005. Global Logistics Indicators, Supply Chain Metrics, and Bilateral Trade Patterns. *SSRN eLibrary*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=869999 [Accessed August 5, 2010].

Institute for Supply Management, 2010. ISM - ISM Reports on Business. Available at: http://www.ism.ws/ISMReport/?navItemNumber=4892 [Accessed September 29, 2010].

Keebler, J.S. et al., 1999. *Keeping Score: Measuring the Business Value of Logistics in the Supply Chain*, Council of Logistics Management.

OPISNET.COM, 2010. Transportation Fuel Index. Available at: http://www.opisnet.com/tfi.asp

[Accessed September 29, 2010].

Ryan Barnes, 2010. Economic Indicators: Purchasing Managers Index (PMI). Available at: http://www.investopedia.com/university/releases/napm.asp [Accessed August 30, 2010].

Shanghai Shipping Exchange, 2009. Shanghai Containerized Freight Index. Available at: http://en.chineseshipping.com.cn/html/scfi.html [Accessed August 26, 2010].

Sheffi, Y., 2004. Combinatorial Auctions in the Procurement of Transportation Services. *Interfaces*, 34(4), 245-252.

The Baltic Exchange, 2010. The Baltic Exchange - Home Page. Available at: http://www.balticexchange.com/ [Accessed September 29, 2010].

The World Bank, 2010. Trade Logistics & Faciliation - Logistics Performance Index (LPI). Available at: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTTRANSPORT/EXTTLF/0,,contentMDK:21514122~menuPK:3875957~pagePK:210058~piPK:210062~theSitePK:515434,00.html [Accessed August 6, 2010].

Tracey, M., 1998. The Importance of Logistics Efficiency to Customer Service and Firm Performance. *The International Journal of Logistics Management*, 9(2), 65-81.

U.S. Census Bureau, 2010. US Census Bureau Manufacturers' Shipments, Inventories, and Orders. Available at: http://www.census.gov/manufacturing/m3/ [Accessed September 20, 2010].

USDA AMS, 2010. Agricultural Marketing Service - Grain Transportation Report (GTR). Available at: http://www.ams.usda.gov/AMSv1.0/ams.fetchTemplateData.do?template=TemplateA&navID=AgriculturalTransportation&leftNav=AgriculturalTransportation&page=ATGrainTransportationReport&description=Grain%20Transportation%20Report%20(GTR)&acct=graintransrpt [Accessed August 24, 2010].

# Appendix B – Commodities Transported

## Commodities Transported

The following Appendix details the commodities used to calculate the aggregated weight transported through each lane. The commodities used are transported using only standardized containers.

### Standardized Containers

1. Cereal grains
2. Agricultural products
3. Animal feed
4. Meat/Seafood
5. Milled grain products
6. Alcoholic beverages
7. Plastics/Rubber
8. Wood products
9. Newsprint/Paper
10. Articles-base metal
11. Machinery
12. Motorized vehicles
13. Furniture
14. Manufactured products
15. Mixed freight

### Non-standardized Containers

1. Live animals/fish
2. Foodstuffs
3. Tobacco products
4. Building stone
5. Natural sands
6. Gravel
7. Nonmetallic minerals
8. Metallic ores
9. Coal
10. Crude petroleum
11. Gasoline
12. Fuel oils
13. Coal-Petroleum products
14. Basic chemicals

15. Pharmaceuticals
16. Fertilizers
17. Chemical products
18. Logs
19. Paper articles
20. Printed products
21. Textiles/Leather
22. Nonmetallic mineral products
23. Base metals
24. Electronics
25. Transport equipment
26. Precision instruments
27. Waste/Scrap
28. Unknown commodities

# Appendix C – Economic Indicators

## Introduction

The following section lists all relevant attributes and indicators of economic performance for lane origins and destinations considered when searching for relevant attributes. All of these attributes were obtained based on the statistical areas of the origin/destination of each specific lane. We relied mostly on Census.gov for data, along with the Federal Highway Administration's Freight Analysis Framework database for several attributes.

## Economic Indicators

1. Population '(000)
2. Rank pop
3. Urban Area Size (square miles)
4. Population Density (persons/sq miles)
5. Peak Period Travelers (000)
6. Freeway Daily Vehicle-Miles of Travel'(000)
7. Freeway Lane-miles
8. Arterial Street Daily Vehicle-Miles of Travel
9. Arterial Street Lane-miles
10. Public Transportation Annual Passenger-miles'(million)
11. Public Transportation Annual Unlinked Passenger Trips
12. Cost Component - Value of Time ($/hour)
13. Cost Components  - Commercial Value of Time'($/hour)
14. Cost Component - Average State Fuel Cost '($/gallon)
15. Congested Travel (% of Peak VMT)
16. Congested System (% of lane-miles)
17. Number of Rush Hours (time when system may be congested)
18. To Maintain Constant Congestion Level Annual lane-miles needed (freeway & arterial)
19. To Maintain Constant Congestion Level Annual Transit or Carpool Riders needed (000)
20. Annual Excess Fuel Consumed Total Gallons (million)
21. Annual Excess Fuel Consumed Rank
22. Annual Excess Fuel Consumed Gallons per Peak Traveler
23. Annual Excess Fuel Consumed Peak Traveler Rank
24. Annual Hours of Delay Total Delay (000)
25. Annual Hours of Delay Rank
26. Annual Hours of Delay Traveler per Peak Traveler

27. Annual Hours of Delay Rank
28. Percent of Delay due to Incidents
29. Travel Time Index Value
30. Travel Time Index Rank
31. Annual Congestion Cost Total Dollars (million)
32. Annual Congestion Cost ACC Rank
33. Annual Congestion Cost ACC per Peak Traveler ($)
34. Annual Congestion Cost per peak travelerACC Rank
35. Annual Effects of Operations Treatments Delay Reduction (1000hours)
36. Annual Effects of Operations Treatments Delay Reduction per Peak Traveler '(hours)
37. Annual Effects of Operations Treatments Wasted Fuel Reduction '(1000 gallons)
38. Annual Effects if Operations Treatments Congestion Cost Savings ($million)
39. Condition if Public Transportation Service were Discontinued Annual Delay Increase '(1000 hours)
40. Condition if Public Transportation Service were Discontinued Annual Effects if Operations Treatments Annual Delay Increase per Peak Traveler (hours)
41. Condition if Public Transportation Service were Discontinued Wasted Fuel Reduction(1000 gallons)
42. Condition if Public Transportation Service were Discontinued Annual Congestion Cost Increase '($million)
43. Roadway Congestion Index
44. Rannally Rating
45. # of Trucking Companies
46. # of Broker Companies
47. Warehouse
48. Regional Diesel Prices
49. Trade surplus or deficiency

## Resources

The Census Bureau conducts more than 100 economic surveys annually. The surveys cover various sectors of the economy, at annual, quarterly and monthly time periods. Data from these surveys is available to the public at American FactFinder.

The Freight Analysis Framework (FAF) incorporates data from multiple sources to create a comprehensive model of freight movement in the United States among major metropolitan area by any mode of transportation. The FAF provides estimates for tonnage and value, by commodity type, mode, origin, and destination using data from 2007 Commodity Flow Survey.

# Appendix D – Lane Clusters

# 1 LANE CLUSTERS

K-means algorithm was used to cluster the lanes in 5 groups. The graphs present the counts for each of the cluster in terms of the regional origin and destination. From these figures, one can observe the general composition of each of these clusters in terms of the regional origin and destinations.

**Lane Clusters**

- Cluster 0: Origin Central Atlantic to balanced destinations—Short distance.
- Cluster 1: Origin/destination Midwest.
- Cluster 2: Balanced origins with destination Gulf Coast.
- Cluster 3: Balanced origins with destination West Coast.
- Cluster 4: Balanced origins with destination Lower Atlantic—Long distance.

## 1.1 CLUSTER: ORIGIN CENTRAL ATLANTIC TO BALANCED DESTINATIONS—SHORT DISTANCE

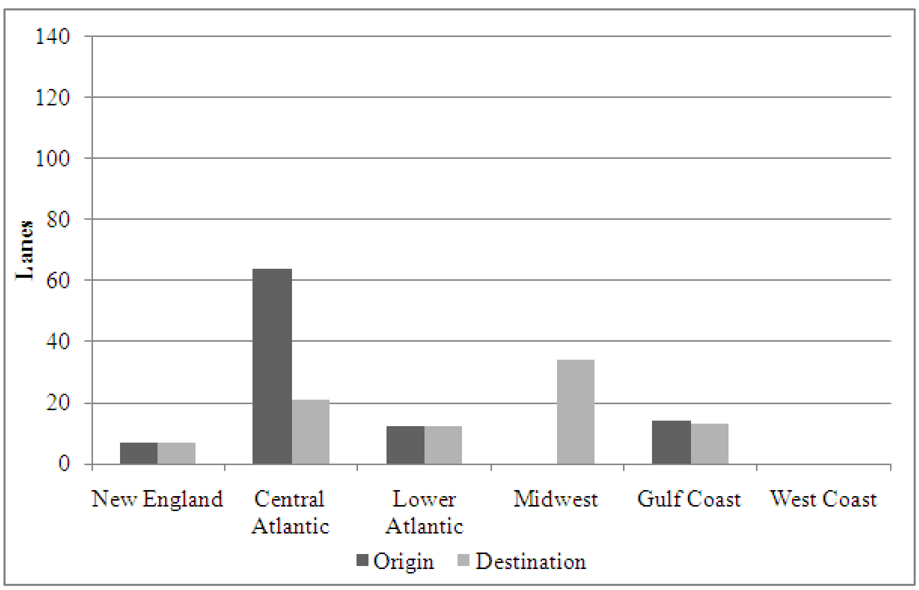


**Figure 1-1 – Distribution of origins and destinations for Cluster 0**

Figure 1-1 presents the composition of cluster 0. One can observe that the destinations are fairly balanced among the different clusters. The regional origin seems to be weighted more towards the central Atlantic and gulf coast areas. Figure 1-2 and Figure 1-3 represent the lanes origins and destinations, respectively, for cluster 0.

**Figure 1-2 – Cluster 0: Origin map**

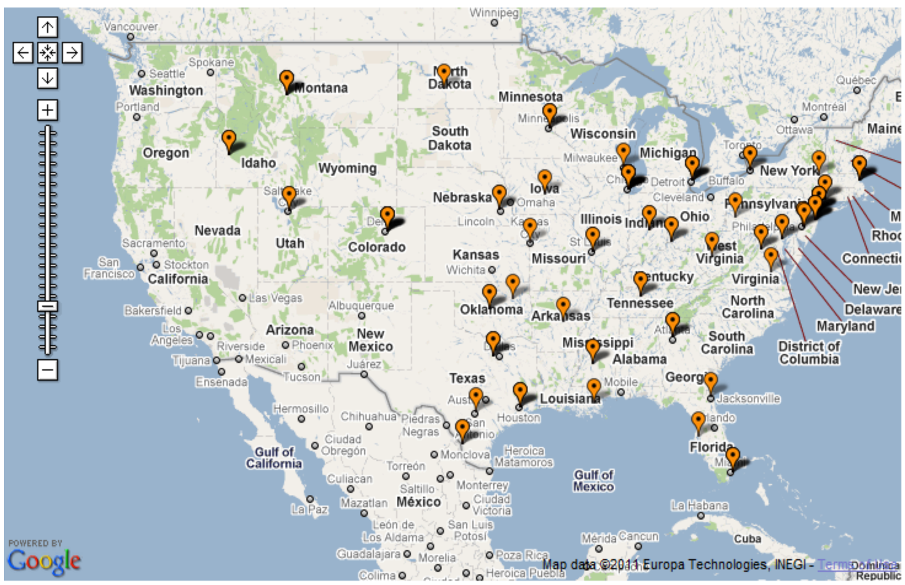

**Figure 1-3- Cluster 0: Destination map**

## 1.2 Cluster: Origin/destination Midwest

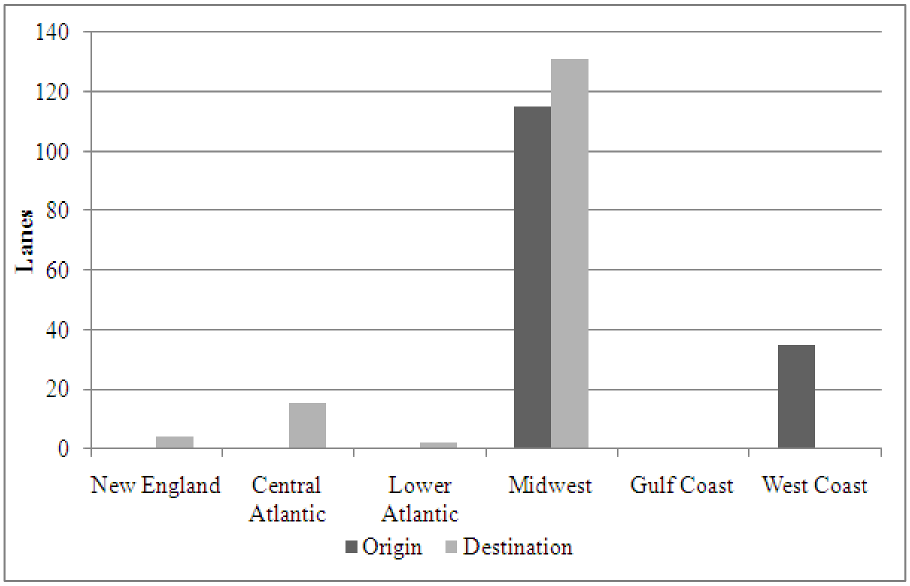


**Figure 1-4 – Distribution of origins and destinations for Cluster 1**

Figure 1-4 describes the composition of cluster 1, in which the Midwest dominates the region of origin and destination.  The west coast region also appears with strong presence as a region of origin.

Figure 1-5 and Figure 1-6 represent the lanes origins and destinations, respectively, for cluster 1.

**Figure 1-5 - Cluster 1: Origin map**

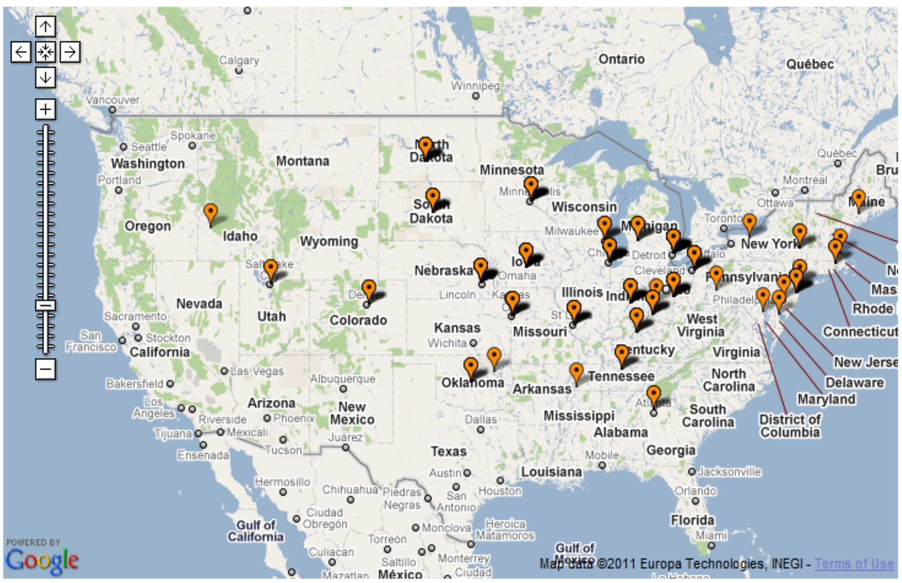

**Figure 1-6 – Cluster 1: Destination map**

## 1.3 CLUSTER: BALANCED ORIGINS WITH DESTINATION GULF COAST

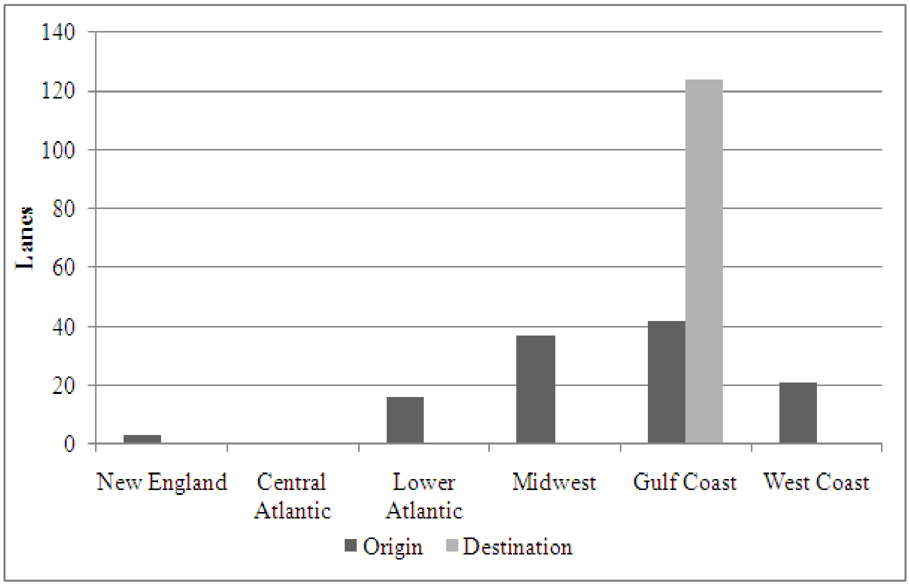


**Figure 1-7 – Distribution of origins and destinations for Cluster 2**

Figure 1-7 presents general characteristics of cluster 2. In this case, one can observe that the gulf coast region has a strong presence as destination. This might be due to the importance of the sea ports in Houston and other areas within the region. The geographic origins of the lanes are distributed among the lower Atlantic, Midwest, gulf coast, and west coast areas. Figure 1-8 and Figure 1-9 represent the lanes origins and destinations, respectively, for cluster 2.


**Figure 1-8 – Cluster 2: Origin map**

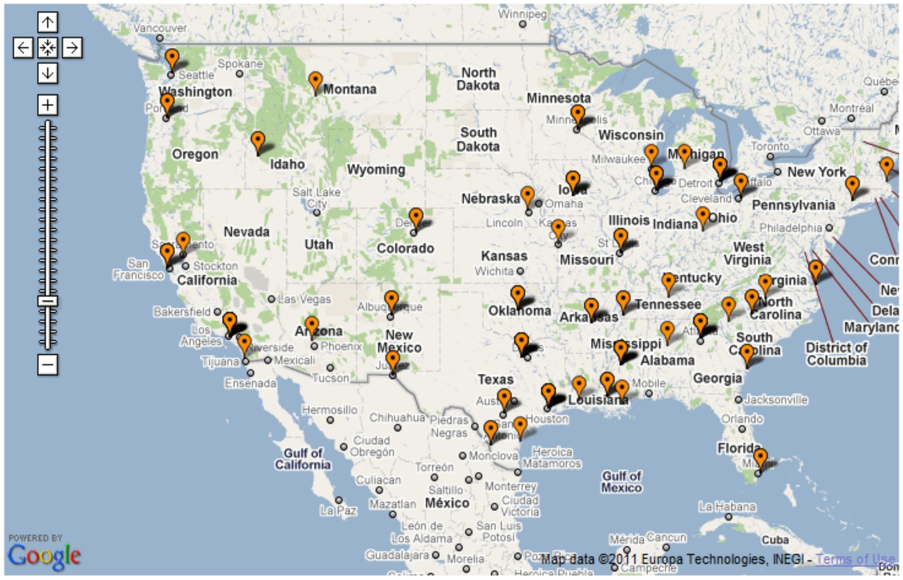


**Figure 1-9 – Cluster 2: Destination map**

## 1.4 CLUSTER: BALANCED ORIGINS WITH DESTINATION WEST COAST

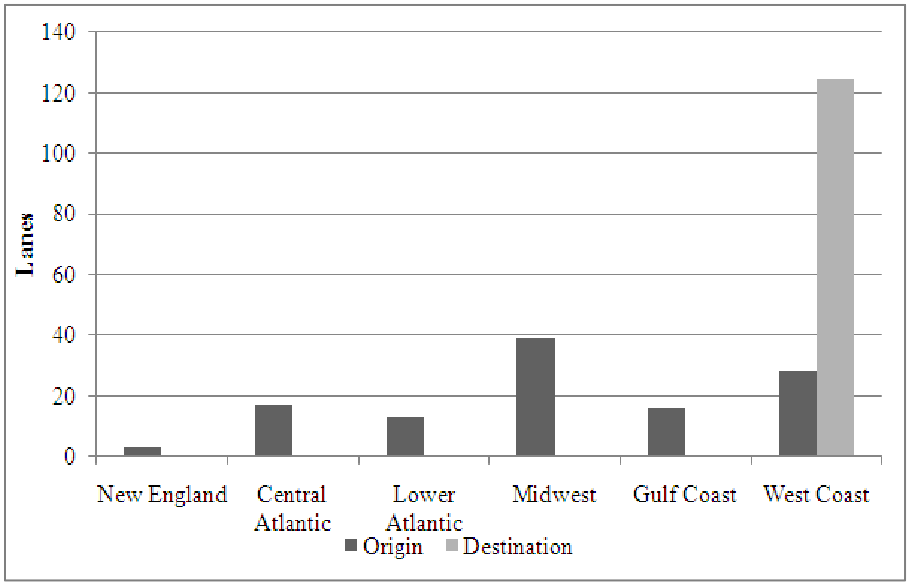


**Figure 1-10 – Distribution of origins and destinations for Cluster 3**

Figure 1-10 presents the general characteristics of cluster 3. In this case, the strong destination presence belongs to the west coast area. This is due to the location of important west coast sea ports such as Los Angeles, Long Beach and Seattle. The origins of these lanes are distributed among the rest of the regions. Figure 1-11 and Figure 1-12 represent the lanes origins and destinations, respectively, for cluster 3.


**Figure 1-11 – Cluster 3: Origin map**

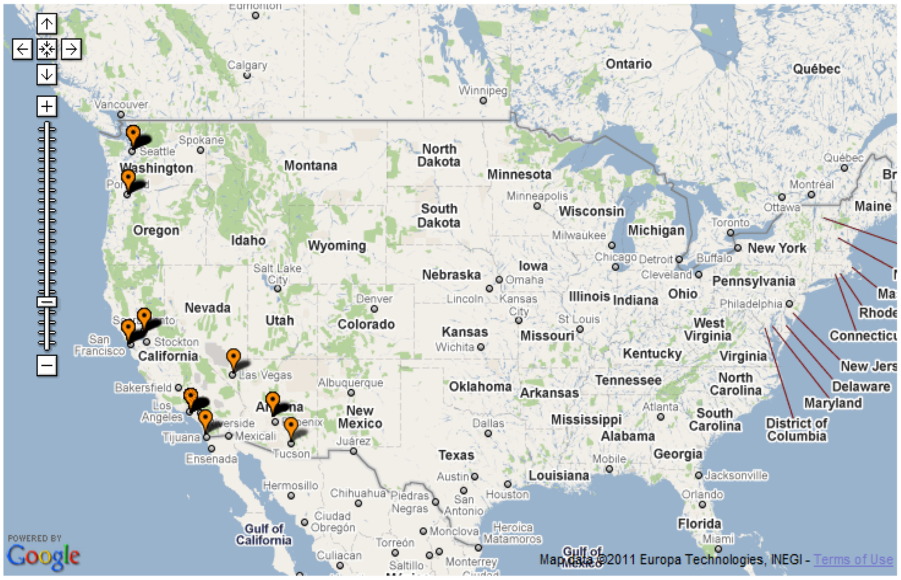


**Figure 1-12 – Cluster 3: Destination map**

## 1.5 Cluster: Balanced origins with destination Lower Atlantic—Long distance

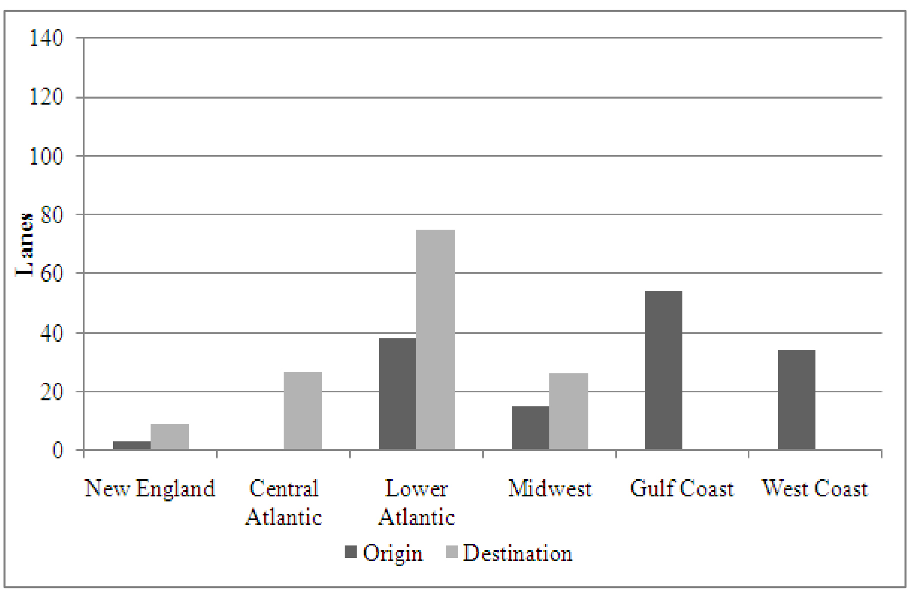


**Figure 1-13 – Distribution of origins and destinations for Cluster 4**

Finally, Figure 1-13 contains the characteristics of cluster 4, whose regional origins and destinations are evenly spread among all the regions. Figure 1-14 and Figure 1-15 represent the lanes origins and destinations, respectively, for cluster 4.

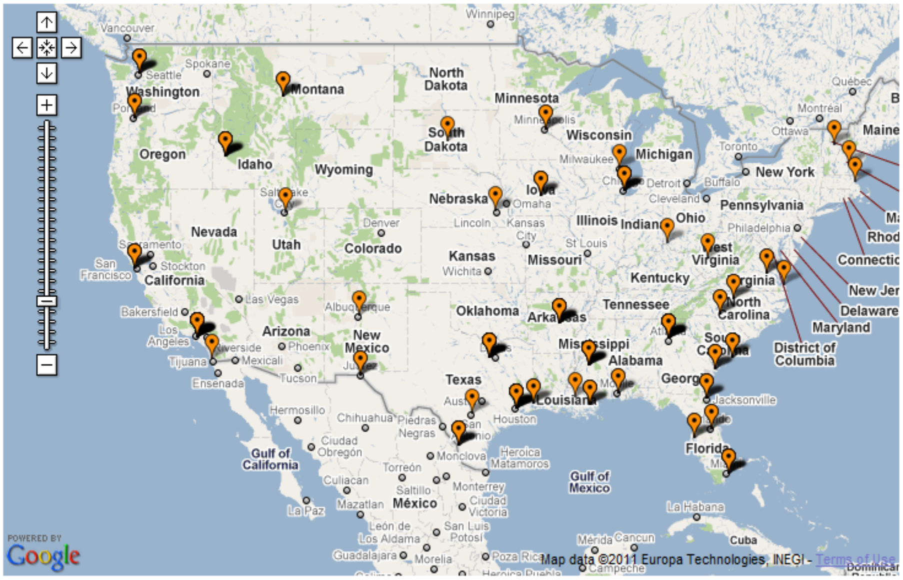

**Figure 1-14 – Cluster 4: Origin map**

**Figure 1-15 – Cluster 4: Destination map**

# Appendix E – DEA Code

## DEA Model code

The following code was used to define the DEA optimization model in AMPL.

```
# Optimization model code for DEA
####################################################################

### Data sets: ###
set LANES ordered;                          # Individual Lanes
set CHARS;                                  # Characteristics of the lane

### Variable and parameter definition: ###
param Data {LANES,CHARS}        >= 0;       # Set of inputs/outputs for the model
param g >=0;
param Efficiencies{LANES}       >= 0;       # Store efficiencies in this vector
var Theta                       >= 0;       # efficiency frontier parameter
var Lam{LANES}          >= 0;               # intensity of firm

## Model: ##

minimize Efficiency: Theta;

subject to Imputs {j in CHARS: j <> "AVGINV" }:
        sum {i in LANES} Lam[i]*Data[i,j] <= Data[g,j]*Theta;

subject to Outputs {j in CHARS: j = "AVGINV" }:
        sum {i in LANES} Lam[i]*Data[i,j] >= Data[g,j];
```

## DEA Input/Output Script

The following code was used to run the DEA model iteratively to calculate the efficiencies of several lanes.

```
# DEA Readscript
#####################################################################

reset;
option solver cplex;

model DEA7\DEA.mod;

#Declaring the data
table LANES IN "ODBC" "DEA7\DEADATA.XLSX": LANES <- [LANES];
table CHARS IN "ODBC" "DEA7\DEADATA.XLSX": CHARS <- [CHARS];
table AllData IN "ODBC" "DEA7\DEADATA.XLSX": [LANES,CHARS], Data;

#Reading the data
read table LANES;
read table CHARS;
read table AllData;

# Definition of Master Problem #
let g := 1;
for {x in LANES} {
let g := x;
solve;
let Efficiencies[x] := Theta;
}

# Write data output #
table EFFICIENCY OUT "ODBC" "DEA7\DEADATA.XLSX": [LANES], Efficiencies;
write table  EFFICIENCY;
```

# Appendix F – Data Structure

## 1 Introduction

The following appendix details the attributes associated with the potential sharing platform used to store and retrieve relevant data. The database developed for handling data throughout the project has several tables with individual attributes regarding each individual lane and other information relating to complementary information such as economic indicators at origin/destination cities or commodities transported. Likewise, the structure is set up to allow for storage of information regarding specific rates for the lanes for any specific date. These attributes were collected from multiple sources throughout the study. This basic data structure could be used for future data collection and analysis; moreover, this basic structure would serve as a base towards developing more comprehensive implementations of the project such as web based applications.

## 2 Database Structure

In Figure 2-1, we detail the structure of the database used. The following tables were created to store all relevant data associated with our study: LANE, ZONE, ZONE_ECON, COMMODITY, DATE, and LANE_SPOT_RATE_HISTORY. All tables contain individual attributes associated with them. These attributes will be further explained in the next section.

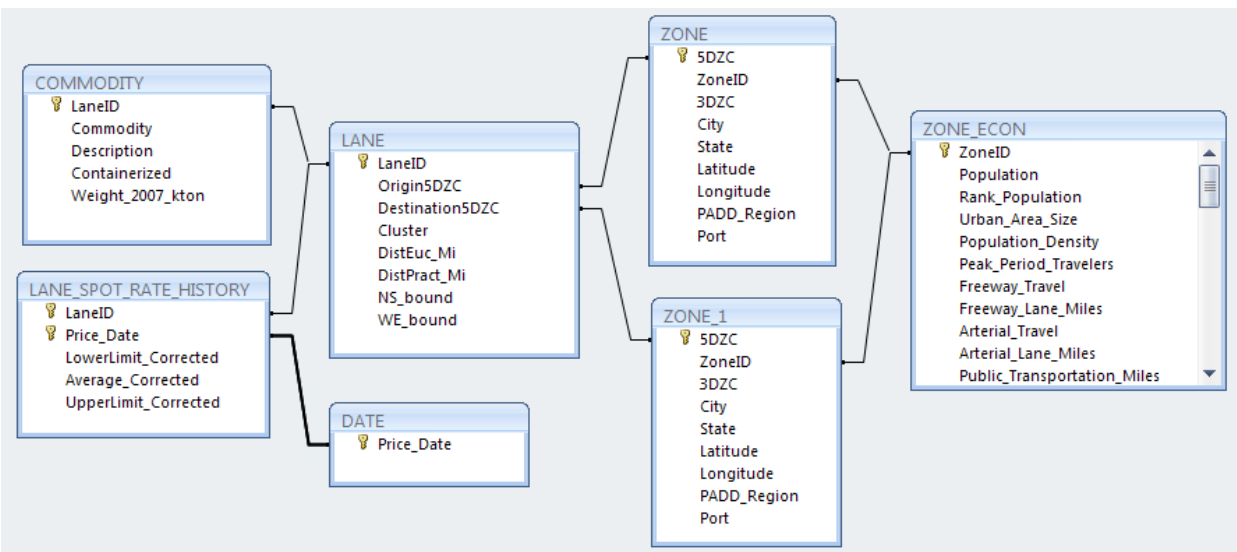


**Figure 2-1– Data Structure**

## 3 Attribute descriptions

LANE table:

- **LaneID** – Is the primary key associated with each lane and the primary unit of measurement in the study
- **Orgin5DZC - ** Origin zip code for each lane
- **Destination5DZC** – Destination zip code for each lane
- **Cluster** – Cluster assignation of each lane
- **DistEuc_Mi** – Euclidian distance between origin and destination in miles
- **DistPract_Mi -** Practical (real) distance between origin and destination in miles
- **NS_bound** – Whether the lane primarily heads South or North
- **WE_bound** – Whether the lane primarily heads West or East

ZONE table:

- **5DZC** – Zip code associated with lane origin and destination
- **ZoneID** – ID given to the US statistical area where the zip code is found
- **3DZC** – Three digit zip code
- **City** – City where zip code is located
- **State** – State where zip code is located
- **Latitude** – Latitude of the zip code under consideration
- **Longitude** - Longitude of the zip code under consideration
- **PADD_Region** – Petroleum Administration and Defense Districts defined regions
- **Port** – Whether there is a port in the statistical area where the zip code is located

ZONE_ECON table: List of Economic Factors, found in Appendix C

COMMODITY table:

- **LaneID**
- **Commodity** – Set of commodities transported in each lane
- **Description** – Description of commodities
- **Containerized-** If the commodity is transported in containers (see Appendix B)
- **Weight_2007_kton**- Weight in Kilo-Tons transported through the lane in 2007

DATE:

- **Price_Date** – Dates ranges corresponding to our data

LANE_SPOT_RATE_HISTORY:

- LaneID
- Price_Date- The specific date for which the data is collected
- LowerLimit_Corrected – The upper limit spot rate as provided by TransCore
- Average_Corrected – The average spot rate as provided by TransCore
- UpperLimit_Corrected – The lower limit spot rate as provided by TransCore