The Impact of Bidding Aggregation Levels on Truckload Rates

by

Julia M. Collins
B.A. Art History and History
Wellesley College, 2005

and

R. Ryan Quinlan
MBA in Management
Brigham Young University, 2006
B.S. Economics
Brigham Young University, 2002

Submitted to the Engineering Systems Division in Partial Fulfillment of the
Requirements for the Degree of

Master of Engineering in Logistics
at the
Massachusetts Institute of Technology
June 2010

Signature of Autho ....
Master of Engineering in Logistics Program, Engineering Systems Division
May 7, 2010

Signature of Author..
Master of Engineering in Logistics Program, Engineering Systems Division
May 7, 2010

Certified by.............................
Dr. Chris Caplice
Executive Director, Center for Transportation and Logistics
Thesis Supervisor

Accepted by..
Prof. Yossi Sheffi
Professor, Engineering Systems Division
Professor, Civil and Environmental Engineering Department
Director, Center for Transportation and Logistics
Director, Engineering Systems Division

The Impact of Bidding Aggregation Levels on Truckload Rates

by

Julia M. Collins and R. Ryan Quinlan

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of
Master of Engineering in Logistics

Abstract

The objective of this thesis was to determine if line-haul rates are impacted by bid type, and if aggregation of bidding lanes can reduce costs for both shippers and carriers. Using regression analysis, we developed a model to isolate and test the cost effects that influence line-haul rate for long-haul shipments. We have determined that aggregation of low-volume lanes from point-to-point lanes to aggregated lanes can provide costs savings when lanes with origins and destinations in close proximity to each other can be bundled. In addition, bidding out region-to-region lanes can supplement point-to-point lanes by reducing the need to turn to the spot market. The model shows that bundling lanes can provide significant cost savings to a shipper because contract lanes of any type are on average less costly than spot moves. This thesis provides guidelines and suggestions for aggregation when creating bids during the first stage of the truckload procurement process.

Thesis Supervisor: Dr. Chris Caplice
Title: Executive Director, Center for Transportation and Logistics

Acknowledgments

I would like to thank our sponsors, Kevin McCarthy, Chris Brady, and Steve Raetz, for their support, good humor, and invaluable guidance in this project.

I would also like to thank our advisor, Chris Caplice. He provided guidance, support, and encouragement throughout the process, and challenged us to do our best.

Thank you to my MLOG classmates for their help and humor all year, especially to my thesis partner, whose sense of humor was invaluable.

And last, but not least, thank you to my family for their love and support throughout my education and especially this year at MIT.

-Julia

My portion of the paper is dedicated to my family. My wife's encouragement and support have made my academic goals possible. Her love and friendship over the last ten years have made my life wonderful.

My children's excitement for life and their understanding of my desire to come back to school are inspiring. They are, and will always be, my greatest achievements.

Lastly, I would like to thank my parents for their support throughout my life and especially during my time at MIT. They continue to be great examples.

-Ryan

Contents

List of Figures

List of Tables

1 Introduction

The US commercial truckload market generates over $320B in revenue (Standard & Poor's, 2010). Truckload rates for shippers are generally procured either annually or bi-annually through a bidding process. While shipments move from a single point to a single point, they can be classified more broadly in the bidding process. For example, a shipment from Boston to Minneapolis can be defined in a bid in many ways, to include: specific address to specific address; by zip code, 02110 to 55421, Boston to Minneapolis; 3-digit zip code, 021 to 554; and Massachusetts to Minnesota. The different classifications reflect the varying geographic definitions that can be assigned to the origin and destination of each shipment.

Our research objective is to determine the impact of various lane definitions on long-haul, dry van, truckload line-haul rates using differing origin and destination sizes. By defining origin and destination sizes more broadly, companies can capture higher shipping density in a lane. However, larger origin and destination regions can lead to higher empty miles following or preceding a load. Our thesis uses linear regression to explore how this additional shipping density affects the price of the shipments in that lane and whether this additional density results in cost savings.

As the size of the origin or destination increases, the total number of shipments moving between those two points will increase because more shipments will be included. Our thesis will examine if the price of the original shipment defined as a lane from Boston to Minneapolis changes when it is defined as a lane from Massachusetts to Minnesota or some region size in between. We will assess how defining lanes differently can impact pricing. Going forward we will refer to this impact as "economies of aggregation". By aggregation we mean combining multiple shipments with somewhat common or close by origins or destination into a single lane for the purpose of collecting a rate driving a bid.

In for-hire truckload (TL) transportation, loads can be bid out using a variety of lane definitions. A shipper could potentially bid out the origin and destination as any geographic area depending on what best suits its needs. However, our data shows that most shippers default to point-to-point bids for their lanes. As will be described in further detail later, our data includes over half a million points. After testing our data we found that it is consistent with general market trends and we will therefore assume that the dataset is a reasonable sample of the U.S. truckload market.

It is unclear if the current bidding strategy provides the best value for shippers and carriers. Our data set shows that most of these point-to-point lanes extremely low volume. Low-volume lanes, with fewer than 15 loads per year, constitute over 85% of the lanes in this study. Could companies use aggregation techniques to save money?

The TL market has three distinctive features that make the economies of aggregation important: cost plus pricing, Economies of Scope, and two-level procurement. As discussed later, TL is a highly competitive market with extremely low margins. Thus, most TL pricing is cost-plus. For example, the line-haul cost for a shipment going from Los Angeles to Salt Lake City is mainly dictated by the distance traveled. The cost is also impacted by the Economies of Scope, or the interdependency between lanes in a carrier network. One implication of these characteristics of the truckload industry is that the line-haul cost for two different shipments will be the same if they originate in the same general area and terminate in the same general area – even if the actual points of pick-up and drop-off are not identical. This phenomenon differentiates trucking procurement from procurement in other industries. Because two different units can act identically in terms of the procurement process, the truckload market offers a unique chance to test how economies of aggregation affect pricing.

The truckload market differs from most other industries' in that buyers use a two-stage procurement process. The first stage is the planning process. This is when companies decide which lanes need to be sent to bid, how to define the lanes, and how long the term of the contract should last. When defining the lanes companies must decide what geographic classifications to use. Although theoretically these geographic classifications could take any form, in our data they are defined using common classifications of warehouse, city, 5-digit zip code, 3-digit zip code, and state. Prices are set in the first stage.

The second stage of the truckload procurement process is the tendering- when an actual shipment is moved. Once rates have been agreed upon in the first stage, companies will tender the truckload services as the need arises. This stage can also influence cost since carriers can decline to ship loads even though there is a contract in place for those loads. Companies can also find that they need to ship loads for which they did not establish contract pricing in the first stage. Loads for which there is no contract rate are bid out on the spot market.

A company that makes truckload shipments can create a procurement package for truckload carriers with shipments defined narrowly (small region to small region), or broadly (large region to large region). Lanes can also be defined as small region to large region and vice versa. The critical question for shippers conducting truckload procurement projects is which route will produce the lowest total cost for the company. While a more broadly defined lane would include more individual shipments, truckload carriers may still choose not to lower rates. Because the region is more broadly defined, carriers may seek to hedge against the uncertainty of where the load will actually end up within the region or state. That is, they may believe that factors associated with a larger region may raise their costs even as the potential for a follow-on load increases.

Figure 1 shows how 5 planned shipments can be aggregated into different bidding lanes. A company that makes truckload shipments can create a procurement package for truckload carriers with lanes defined narrowly or broadly. For example Figure 1b creates a state to state lane where all 5 shipments would move under the same rate. While a more broadly defined lane would include more individual shipments, truckload carriers may still choose not to lower rates. Figures 1c and 1b use multiple levels of aggregation. For example, if the Chicago area has a higher number of available shipments then carriers may be willing to accept a larger aggregated area since they would expect to be able to find another load relatively easily. If this is the case then bundling the Chicago area might capture the aggregation possibilities in the Chicago region. (See figures 1c and 1d.)

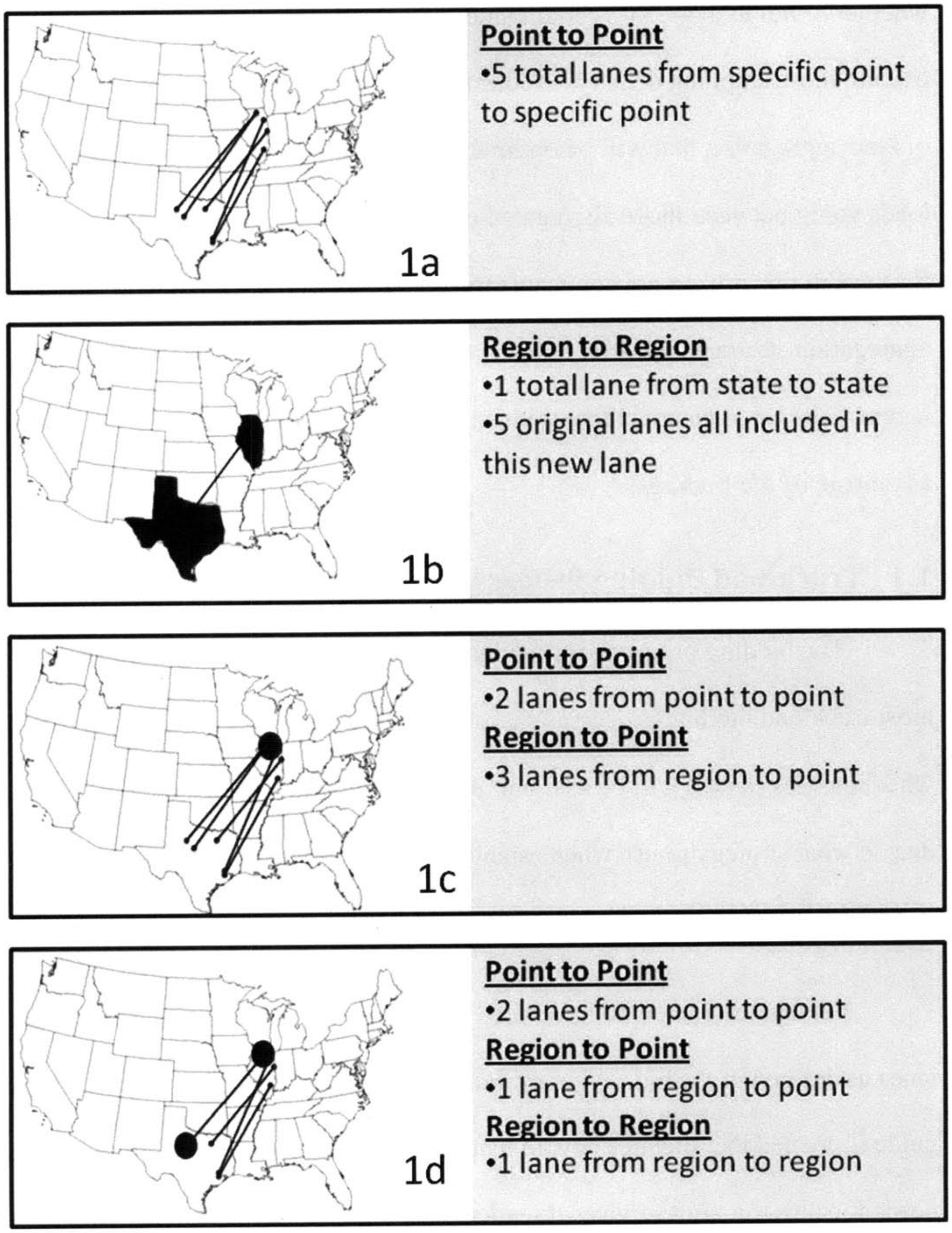


Figure 1: Example of Different Lane Aggregation Levels

Our research has broad applicability for companies that utilize for-hire truckload transportation. Companies can use this research to develop their strategy for procuring truckload rates. When reviewing truckload shipping history, companies can decide whether or not to aggregate certain lanes into larger geographic routes or lanes in order to capture more shipping density. Additionally, we will provide insights regarding the level of lane aggregation that will be optimal for such companies and how the total number of loads going between those aggregated points can influence the per unit cost. Our findings in this project are applicable to companies who have implemented a bidding aggregation strategy and want to evaluate their results against the findings contained herein. The research concludes with suggestions for companies who wish to take advantage of the findings.

1.1 Truckload Bidding Process

The bidding process for truckload transportation varies by company. However, most truckload bidding processes share a common structure. Companies first decide what business (lanes) will be sent out to bid and how to package the bid. They must also decide what strategy to use when establishing the shipper-carrier relationship (Caplice and Sheffi, 2003).

For-hire carriers who have been invited to participate then submit pricing for the lanes as defined in the bid. After evaluating and analyzing the submitted bids, a routing guide is created that dictates how to handle each of the lanes in the bid. As the company needs for-hire trucking services it tenders the shipments to carriers based upon the established routing guide. The routing guide specifies a primary and alternate carriers by lane. If the carriers on a routing guide do not accept a load then the company must get a "spot quote."

1.2 Motivation

Shippers trying to build an effective truckload transportation program must decide which carriers to use and how to build an optimal routing guide. Because routing guides typically remain in effect for one to two years, shippers and carriers are making a long-term commitment to the rates contracted. Our data shows that line-haul rates paid for long-haul truckload shipments are mostly based on origin and destination and line-haul distance. In fact, total distance and location explain 90% of the variability of line-haul cost in our model! The remaining 10% of the variability is where shippers have the opportunity to capture cost savings. Therefore, understanding how economies of aggregation affect truckload pricing is relevant to any company who routinely procures truckload services.

1.3 Structure

This paper is organized as follows: Chapter 2 provides an introduction to the for-hire truckload trucking industry. Chapter 3 discusses the existing body of literature on the topics of truckload trucking, for-hire trucking auctions, and complementary analysis to that which this thesis explores. Chapter 4 outlines the methodology we used and describes the data. Chapter 5 describes the analysis we performed to understand the data and the model we developed to test aggregation levels. Chapter 6 offers the conclusions that can be drawn from our model and analysis.

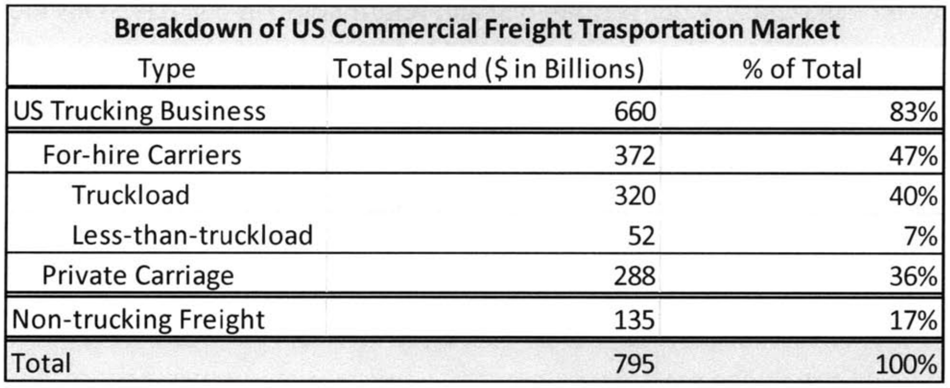
2 Background: Shippers and the Spot Market

Our thesis is specifically focused on dry-van, long-haul truckloads. Dry-van truckloads are those that do not need refrigeration and use a traditional van unit attached to a tractor engine. We define "long-haul" as any load greater than 250 miles from origin to destination. Any load less than 250 miles is generally not considered a long-haul load since a carrier can typically deliver the load and return home in one day. Our thesis will not include any analysis of other types of truckload shipping, such as refrigerated, flatbeds, curtain vans, or intermodal. Short-haul trucking is driven by more regional influences that will not be explored here. It is hoped that the same techniques developed in this thesis could provide a foundation for others seeking to extrapolate the findings here to specialty truckload markets.

Because this research deals specifically with bidding strategy surrounding volume aggregation, it is also important to understand how trucking services are procured and rates are established. Below is a brief introduction to these elements.

2.1 For-Hire Truckload Shipping

The trucking industry is both large and fragmented. Trucking represents 83% of the United States commercial freight transportation market (S&P, 2010). The high level of fragmentation and lack of dominant industry players is due in part to low barriers to entry and exit (IBIS World, 2010). For example, the trucking industry is made up of both private carriage and for-hire trucking. The for-hire sector includes both truckload (TL) and less-than-truckload (LTL) carriers. The American Trucking Association estimates the private carriage sector to account for $288 billion in 2008 (S&P, 2010). Outside organizations contract with for-hire carriers to move freight at a pre-determined rate. Truckload freight typically moves in loads greater than 10,000 pounds, which is the breakpoint where the variable costs of LTL shipments exceed the fixed costs of TL shipment (Mulqueen, 2006). This thesis project focuses solely on for-hire truckload carriers.



Breakdown of US Commercial Freight Trasportation Market

Type	Total Spend ($ in Billions)	% of Total
US Trucking Business	660	83%
For-hire Carriers	372	47%
Truckload	320	40%
Less-than-truckload	52	7%
Private Carriage	288	36%
Non-trucking Freight	135	17%
Total	795	100%

Table 1: US Trucking Market based upon Survey by Standard & Poor's (S&P, 2010)

S&P's survey also reported that in 2008 for-hire carriers generated $372 billion in revenues. Therefore, these revenues make up 56% of the motor carrier market. $320 billion of this business was generated from the TL shipments and the remaining portion from LTL.

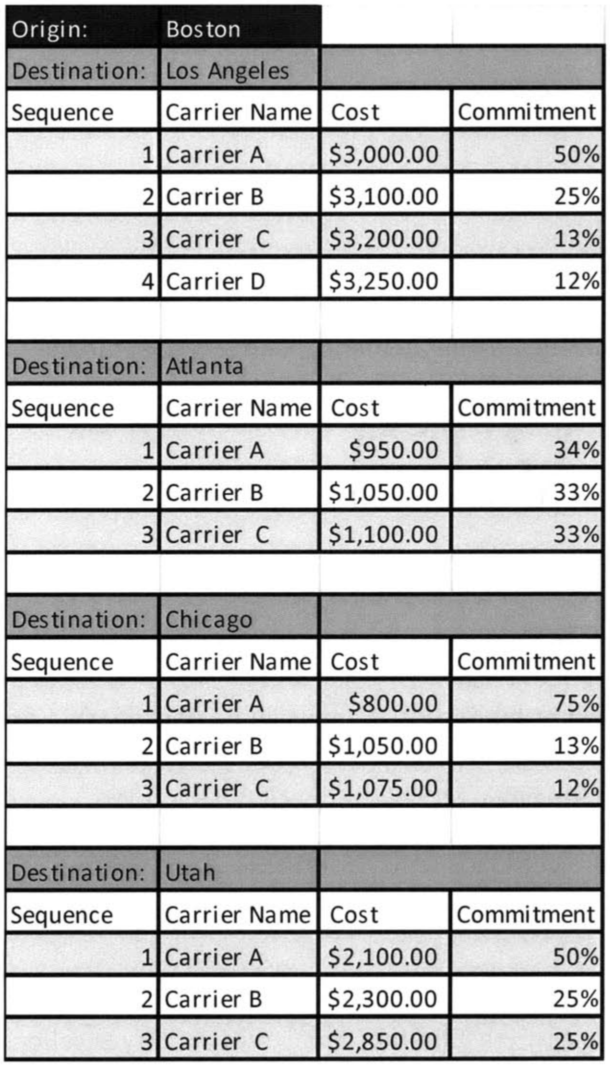
The for-hire TL market is dominated by private carriers. "The industry is dominated by owner-operators, with 88.0% of all establishments being non-employers" (IBISWorld, 2010). This means that the vast majority of players in the for-hire trucking industry are owner-operators, who are often hired as subcontractors by larger carriers, such as J.B. Hunt. Carriers operating more than one vehicle account for only 12% of the total market. S&P reports that the largest publically traded carrier in 2008 was J.B. Hunt Transport Services, with revenues of $3.7 billion in 2008. The largest private TL carrier was Schneider National Inc. which also had $3.7 billion in 2008 revenues. Combined, these two carriers only account for a little over 2% of the U.S. for-hire TL market in 2008.

2.2 Contract Lanes and Routing

To ensure that there will be carriers available for their shipments, shippers typically contract with multiple for-hire carriers, and they set rates for specific lanes. These will be referred to going forward as contract lanes. If a shipper wishes to ship something on a lane for which there is no contract, they can solicit a spot quote from carriers. This rate is valid only for a specific shipment at a specific time. It does not become part of the contracted rate structure for the shipper. Both shippers and carriers can be reluctant to contract lanes for which there will likely be low volume. Contracted lanes are often not collectively exhaustive, so shipments that fall outside of the route guide must be bid out on the spot market.

Contract lane rates are usually for the line-haul, that is, the travel from origin door to destination door. The miles that the carrier has to drive with an empty truck ("empty miles") either to pick up the load or after delivering the load, are not the responsibility of the shipper. However, when pricing a lane, a carrier will factor estimated empty miles into the bid. Carriers try to maximize their fleet's utilization and therefore seek to minimize the total empty miles traveled. Vehicle utilization is measured by the ratio of $\frac{Loaded\ Miles}{Total\ Miles}$ (Standard & Poor's, 2010). As the geographic area of the origin, destination, or both, increases for a given lane, the carriers will try to estimate the average empty miles required to service the load. Loads that require a carrier to travel significant distances at the origin or destination are commonly understood to result in a carrier quoting a higher rate or declining a load (Mulqueen, 2006). The key question of this thesis is whether aggregation can optimize the costs of driving more empty miles and give carriers the benefits of more lane volume. While more volume can make a lane more profitable to a carrier, driving more empty miles due to a broadly defined origin or destination can also offset profitability.

Table 2 below represents a hypothetical routing guide. Assuming that this routing guide is complete for all shipments originating in Boston, this particular company would have four contract lanes, Boston to Los Angeles, Boston to Atlanta, Boston to Chicago, and Boston to the state of Utah. For each one of the lanes there is a list of carriers under contract rates. This hypothetical routing guide also includes price and commitment level between shipper and carrier. Any truckload shipment not explicitly defined in this routing guide would require a spot quote. For example, a shipment originating in Boston and going to Dallas would not have a contract rate in place and would need a spot rate. Furthermore, a shipment going from Boston to Thousand Oaks, CA would also require a spot rate despite its proximity to the Los Angeles market.



Origin:	Boston		
Destination:	Los Angeles		
Sequence	Carrier Name	Cost	Commitment
1	Carrier A	$3,000.00	50%
2	Carrier B	$3,100.00	25%
3	Carrier C	$3,200.00	13%
4	Carrier D	$3,250.00	12%
Destination:	Atlanta		
Sequence	Carrier Name	Cost	Commitment
1	Carrier A	$950.00	34%
2	Carrier B	$1,050.00	33%
3	Carrier C	$1,100.00	33%
Destination:	Chicago		
Sequence	Carrier Name	Cost	Commitment
1	Carrier A	$800.00	75%
2	Carrier B	$1,050.00	13%
3	Carrier C	$1,075.00	12%
Destination:	Utah		
Sequence	Carrier Name	Cost	Commitment
1	Carrier A	$2,100.00	50%
2	Carrier B	$2,300.00	25%
3	Carrier C	$2,850.00	25%

Table 2: Example of hypothetical routing guide for points and region

2.3 Capacity Commitments and Load Rejection

A typical routing guide might also contain a capacity commitment as shown in Table 2 above. A capacity commitment requires carriers to haul a defined percentage of loads over the contracted period. This allows the shipper to maintain flexibility to handle the variation in actual demand and commits the carrier to making the established resources available for each lane. However, according to our contact at a major third party logistics provider, even when capacity commitments exist, they are often not strictly observed, by either the shipper or carrier. Capacity commitments provide more of a guideline of expected volume rather than a strict requirement for either party.

Capacity commitments are necessary and important in part at least because for-hire carriers have the ability to decline loads as mentioned earlier. Capacity commitments aid shippers in designing their networks and forecasting demand. Carriers reject loads based upon limited capacity or because the carrier simply does not want to haul that particular load, typically because it is not profitable for the for-hire carrier at a specific point in time (Mulqueen, 2006). Due to shifting market conditions and changes within individual companies, certain lanes can become unprofitable to carriers even after a contract between shipper and carrier has been established. A lane can be unprofitable for a carrier because of insufficient volume on the lane. Given our thesis focus, it is important to understand carriers' ability to reject loads and the influence lane volume can have on those decisions.

Rejection frequency increases as the carriers gain more power in the buyer/seller relationship (Mulqueen, 2006). Because the balance of power can shift between carriers and shippers depending on market conditions, the frequency of load rejection can also vary depending on carriers' need for revenue. Harding (2005) estimated the cost of rejecting a load to be between 2% and 7% of freight costs. Over 25% of the loads in Harding's research were rejected by the primary carrier (Harding 2005). Over the two years considered in this study, the rejection rate was 26%, with 32% of loads rejected in 2008 and 19% rejected in 2009. Possibly the reason for the rejection rate reduction in 2009 was due to general market conditions. Compared with 2005, 2009's economy was weaker. Presumably for-hire carriers would be less willing to reject a load while market conditions were unfavorable. With a quarter of the loads tendered being rejected, and lane volume contributing to this result, the importance of understanding how to aggregate low-volume lanes becomes extremely relevant to shipping companies.

2.4 Further Explanation of Spot Market

Spot bids play an important part in this research and in the stability of contract lanes. Having contract lanes is typically preferred to exclusively using spot bids due to the difficulty in consistently finding trucks and uncertainty of price. As stated above, shippers must obtain a spot quote for every individual shipment if no contract rate is in place. Depending on the company's system for soliciting requests for spot quotes in the market, this process can take differing amounts of time and effort. Prices in the spot market can also vary depending on the lane, time of year, availability of carrier, and many other factors. Even if a company frequently bids loads in the spot market and is therefore able to predict price, choosing to use the spot market for truckload inherently has more risk in price variability than using a contract rate. However, companies often rely on spot quotes to assess whether its current contract rates are still competitive in the market or if there is reason to renegotiate rates. For example, referring to Table 2, if this company wanted to check the competitiveness of its Boston to Atlanta rates it may seek spot bids to ensure that the three contract lanes it has in its routing guide are still competitive in the market place. This exercise is not necessarily especially practical, however, because a carrier has little impetus to go outside the routing guide, which may be over 20 carriers deep for a particular lane. This study will evaluate whether spot bids consistently influence line-haul rates.

2.5 Economies of Scale, Scope, and Aggregation

Economies of scale define the impact volume has on average unit price. In TL transportation, economies of scale claim that the addition of volume on a lane will lower the average line-haul unit price. If more volume were added to a Boston to Dallas lane, economies of scale predict that the average unit price for the line-haul portion would decrease. While our thesis deals directly with volume on lanes, it is not focused on the effect of adding or subtracting volume from existing lanes. Rather, our focus will deal with the impact of redefining lane boundaries so that more loads are included in a lane. Specifically, aggregating low-volume lanes into newly defined lanes will be analyzed.

Economies of scope hinge on interdependency between lanes in a carrier network. Carriers face uncertainty in securing a follow-on load, both in terms of distance travelled to the follow-on and time spent securing it. The costs of time and distance between loads are built into the tender rate that a carrier generates. When a carrier is better able to predict the probability of a follow-on load, the carrier can also better estimate the costs of interdependency. The greater the probability of a follow-on load, the lower the cost to the carrier and the shipper. Consequently, when a carrier's network can be optimized, it can realize economies of scope (Caplice 1996). For example, a network that covers loads moving from point A to point B would likely be positively influenced by the addition of loads moving from point B or to point A. Truckload pricing is greatly influenced by the "interdependency problem" (Caplice and Sheffi, 2003) due to economies of scope. Truckload carriers who are unsure of the availability of backhaul loads may hedge their pricing to account for this uncertainty. While we will explore the effects of empty miles travelled at the origin and destination, we will not explicitly be exploring the effects of economies of scope.

This thesis examines how the cost of a lane defined as Boston to Dallas compares to a broader lane defined as Massachusetts to 3-digit zip code 752 (which includes Dallas). Practically speaking, it is assumed that the economies of scope that are embedded within the truckload transportation network deeply influence TL rates. While our model does not focus on economies of scope they are generally measured using regional values at the state, 3 digit zip, or other levels. This effect will be discussed later in more depth.

We will focus solely on the impact lane aggregation has on TL rates due to economies of aggregation. The focus of this thesis is on the effects of aggregating bidding lanes to form larger shipping corridors. Economies of aggregation imply that as the size of origins or destinations of a lane decreases, the unit price for that specific lane will decrease. Aggregation does not deal with the addition of new shipments to a lane. Rather it involves the reengineering of lane boundaries at the origin and the destination to include more loads that already exist in a company's portfolio of shipments.

This paper will focus more on the impact of aggregation on pricing and will seek to isolate those effects relative to line-haul rates. However, a complete explanation of influencing drivers to line-haul truckload pricing should include careful analysis of the impact of backhaul availability.

2.6 Low-volume Lanes for Shippers

Low-volume lanes create a problem for both shippers and carriers. Shippers are unsure of how best to bid out these lanes since they typically result in higher costs. Carriers that contract low-volume lanes can have difficulty fitting these lanes into their existing networks. If carriers are unable to find a backhaul (the connection between the end of one load and the beginning of the follow-on load), they risk having to drive the full length of haul without a revenue-generating load at worst, or traveling a long distance for a load. This results in wasted time, labor, and money. Shippers can be penalized in their contracts for these lanes, paying a premium to offset the carriers' extra costs. While the time cost will vary from company to company, the lack of contract line-haul rates can be partly quantified by any premium paid in spot rates. Our data shows that there is a premium for using the spot market instead of establishing contract rates. Aggregation could be a solution to paying this premium.

If shippers can find a way to increase carriers' certainty on low-volume lanes, then this could lead to lower rates for the shipper and better network optimization for the carriers. One possible method to achieve this is to bundle low-volume lanes, defining the lane by using larger origin and destination areas, creating a more collectively exhaustive set of contracted rates. We will address whether aggregating low-volume lanes together will be viewed any differently by carriers than a collection of point-to-point lanes.

Companies seeking to understand how to decrease transportation costs by aggregating low-volume lanes into specific bundles should understand the relationship between the quantity of loads on a lane, the total distance of the lane, and the expected area the carrier will drive empty to pick up or deliver the load. Understanding the relationship between these variables will help to formulate a bidding strategy that will control the total cost of truckload transportation while maximizing the efforts of the employees responsible for running the operation.

3 Literature Review

We conducted a review of the literature related to truckload bidding processes. We also interviewed industry experts who added key insights regarding their observations regarding the impact of region size on line-haul rates. This section reviews our research findings and the relevance of our paper's findings in the field of study.

There was no specific research about the impact of region size and aggregation for a lane on truckload rates. However, there is a good deal of material regarding combinatorial auctions, impact of backhauls on contract truckload rates, and industry practices and standards for truckload rates.

Our literature review is comprised of four categories: literature that contextualizes our project in the broader TL industry; existing efforts to optimize a shipping network; research that took a similar approach to a truckload bidding problem; and truckload carrier economics.

3.1 Industry Context

A significant body of literature exists that characterizes the truckload industry. Caplice and Sheffi (2006) provided a thorough grounding in current bidding practices. Hubbard (2001) explains market factors that are very relevant to our research. The thinner the depth of the local market, the greater the chance of a spot bid on long haul routes. That is, the fewer carriers that operate along a given lane, the less depth there can be in the routing guide. This relates to problems with finding backhauls. This effect is not present when local markets are thick. It provides a clear explanation of the factors affecting long-haul shipping. A decentralized market and less than ideal information about short term supply and demand for a given route are key factors affecting a carrier's experience in a given market. To coordinate backhauls for outbound freight requires a substantial time investment, about 60 man-hours when markets are thick, and two to four times that time when markets are thin (372). This research has helped us characterize some of the factors affecting low-volume lane pricing. Hubbard also contextualizes our research by helping us to understand the factors that affect carrier decisions in low-volume markets.

The ratio of inbound to outbound shipments has a strong impact on common carriage use- the relationship between common carriage use and ratio of inbound to outbound is positively correlated and becomes stronger as the ratio increases (384). Hubbard's findings support our hypothesis that the probability of a backhaul is a significant factor in lane pricing, and would provide an incentive to aggregation.

3.2 Existing Research

The existing body of research about optimization focuses on combinatorial auctions. Combinatorial auctions are a type of bidding where the goal is to optimize the carrier's network. Traditionally, shippers accepted or rejected carriers' bids on a lane-by-lane basis. This made it very difficult for carriers to optimize their networks. Carriers aim for continuous flow within their networks. That is, they would like their drivers to be able to move from point A to point B to point C back to point A without significant detours or interruptions. Carriers can optimize their networks across multiple shippers. However, combinatorial auctions make creating an optimal network easier for carriers because shippers bid on groups of lanes rather than individual lanes. Caplice (2006) provides an overview of this process. This describes the electronic bidding process, as well as current practices in for-hire Truckload shipping. Combinatorial auctions occur later in the bidding process than our research. Combinatorial auctions bundle lanes to increase acceptance by shippers. Our research focuses on creating more optimized lanes which will in turn give better inputs to combinatorial auctions. We also focus more on the potential benefits to shippers rather than carriers. Even though we do not analyze package bids, how carriers respond to lane bundling is interesting in understanding the impact of economies of aggregation.

Sheffi (2004) provides a clear description of the optimization problem that carriers must solve in creating bids for combinatorial auctions and how the transportation services procurement process functions. It also explains clearly the ways that efficiencies can be achieved in the network. Economies of scale can be difficult to achieve, but realizing economies of scope can lead to substantial cost savings. Combinatorial auctions have been a focus for achieving these economies of scope, and we believe similar economies of aggregation can be achieved by optimizing lanes.

Mulqueen (2006) focuses on identifying how to choose between using contract carriers versus using a private fleet. Some of the key variables in deciding whether to privatize a fleet are corridor volume and load variability. In doing the analysis, Mulqueen finds virtually no correlation between high lane volume and lower rates. He could not assess in his project whether there is a correlation between overall shipper TL volume and lower prices from carriers. Mulqueen did find that shippers give significant discounts on lanes that terminate in a known load origination point. That is, carriers will discount freight if they know that the destination of the load is headed directly to another shipping facility from which that carrier will haul another load. Even if the carrier only had to travel an additional 5% of the distance of the first load away from the destination point, the difference in cost was significant (45-47). Our analysis is solely on contract carriers and how to address low-volume lanes, but Mulqueen's research provides a baseline understanding of correlation between lane volume and price.

Zhelev (2004) provides a model for estimating contract rates vs. average spot rates for his data sample. He also finds that spot rates can be as much as 50-100% higher than contract rates. He states that spot rates are 20% higher than contract rates. Zhelev finds that the underlying factor that drives savings in TL transportation is the distribution of loads throughout the year (47). The relationship between contract and spot rates is very important to our analysis. For customers who have a defined transportation strategy in place, spot lanes are inherently low volume, as they were not included in the usually extensive bidding process that shippers undertake. Spot lanes are typically point-to-point; they represent disaggregation in the lane definition continuum. Since we are evaluating how best to aggregate low-volume lanes, understanding the cost tradeoffs between spot lanes and contracted lanes is critical to our analysis.

Harding (2005) provides a method for determining the amount of additional freight expenditures as a result of the frequency and severity of unplanned freight. This contributed to our overall understanding of the cost factors that affect trucking costs. One of the goals of this paper is to enable shippers and carriers to reduce the amount of unplanned freight by making lanes more efficient and more collectively exhaustive.

3.3 Similar Analysis of Truckload Procurement

Caldwell and Fisher's (2008) study of the impact of lead time on truckload transportation rates provided a model to approach the problem of assessing the impact of bidding elements on truckload rates. While their thesis examined lead time's impact on rates, we will use the same methodology to study the impact of lane aggregation on rates. Their work contributed a great deal to our understanding of the data structure. Like this project, theirs was a procurement problem, and consequently involved similar methodology for the model's foundation. The model in this project reflects similar underlying assumptions about which elements most affect truckload line-haul pricing in the United States.

3.4 Truckload Carrier Economics

Caplice and Sheffi (2003) explain the impact of backhaul loads on truckload pricing. When carriers create a network that utilizes backhauls to drive down costs, they are leveraging economies of scope rather than scale. When a carrier is unable to create a network that accurately estimates the connection costs, that is, when they are unable to create economies of scope with certainty, it can hedge its potential costs by increasing a TL carrier's price (Caplice and Sheffi, 112).

3.5 Foundations for Economies of Aggregation

Our review of the literature outlined above led to the following key insights. These insights are helpful to our understanding of the impact of economies of aggregation. Our thesis will fill a gap in existing research by evaluating an earlier stage of the procurement process, particularly about the correlation between lane volume and line-haul rates. Existing research enhanced our understanding of the TL, for-hire market and helped us to better frame our objectives, particularly understanding how our research would enhance current optimization methods by creating a higher quality input.

4 Methodology and Description of Data

In order to investigate the impact of aggregation on total truckload transportation costs, we used two years of transaction data from the Transportation Management Center (TMC), a fee-based, staffed Transportation Management Service, at C.H. Robinson (CHR), a major third-party logistics provider (3PL). The data contains customers comprised of dry-van, truckload shipments for the years 2008 and 2009 originating from the 48 contiguous United States and Washington, D.C. with destinations in the 50 states and Canada from eight different firms using a common TMS. We analyzed long-haul shipments within the 48 contiguous states and Washington, D.C.

Specifically we excluded loads from the dataset with any of the following characteristics:

- Origin or destination outside of the continental US
- Length of haul < 250 miles
- TMC customers with fewer than 2,000 loads over the two year period
- Less-than-truckload shipments
- Rate less than \$0.70 per mile
- Rate greater than \$3.50 per mile
- Blank entries or loads with other obvious data entry errors

After the data had been cleaned, we were left with 495,394 loads out of our original set of 1,101,590. The customer count was reduced from 23 to eight. Between them, the shippers use a total of 618 carriers.

4.1 Data Profiling

To ensure that the data was a representative sample, we conducted various analyses after preparing the data. We wanted to understand how the data was distributed and identify basic characteristics of the customers.

Since we are addressing whether or not to aggregate lanes, we are interested in knowing how the shipments are distributed across lanes. Figure 2 shows that 85.5% of the loads are distributed across lanes with 15 loads per year or less. Only slightly more than 3% of the lanes have more than 150 loads per year. This data includes all of the carriers in the system and all of the customers. Therefore, this analysis represents trends in our data as a whole and is not unique to specific shipping companies or customers.

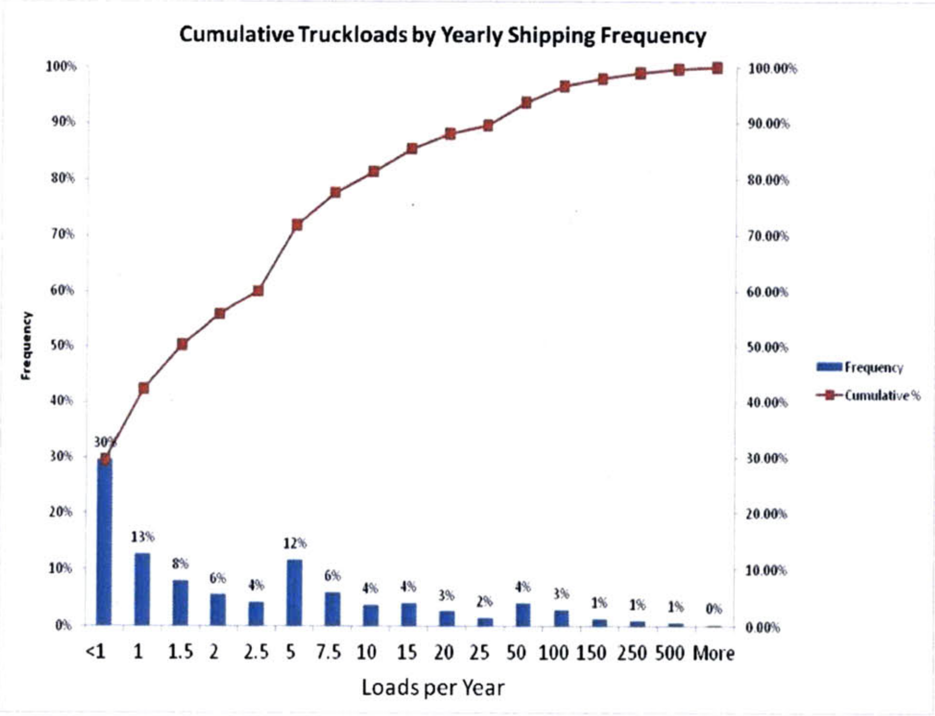


Figure 2: Distribution of Lane Volume

Figure 3 shows the distribution of length of haul by lane. It appears that the length of haul follows a truncated normal distribution. However, since loads under 250 miles and over 3,000 miles were dropped from the final data set, the lower and upper bounds of the dataset were artificially imposed. The length of haul ranges from 250 miles to 2,998 miles. About two thirds of the loads are between 282 and 1,032 miles, within one standard deviation of the mean of 657 miles. That is, of the 495,394 total loads, about 336,868 fall in this range, with almost all of the remaining 158,526 loads falling between 1,1033 and 2,998 miles.

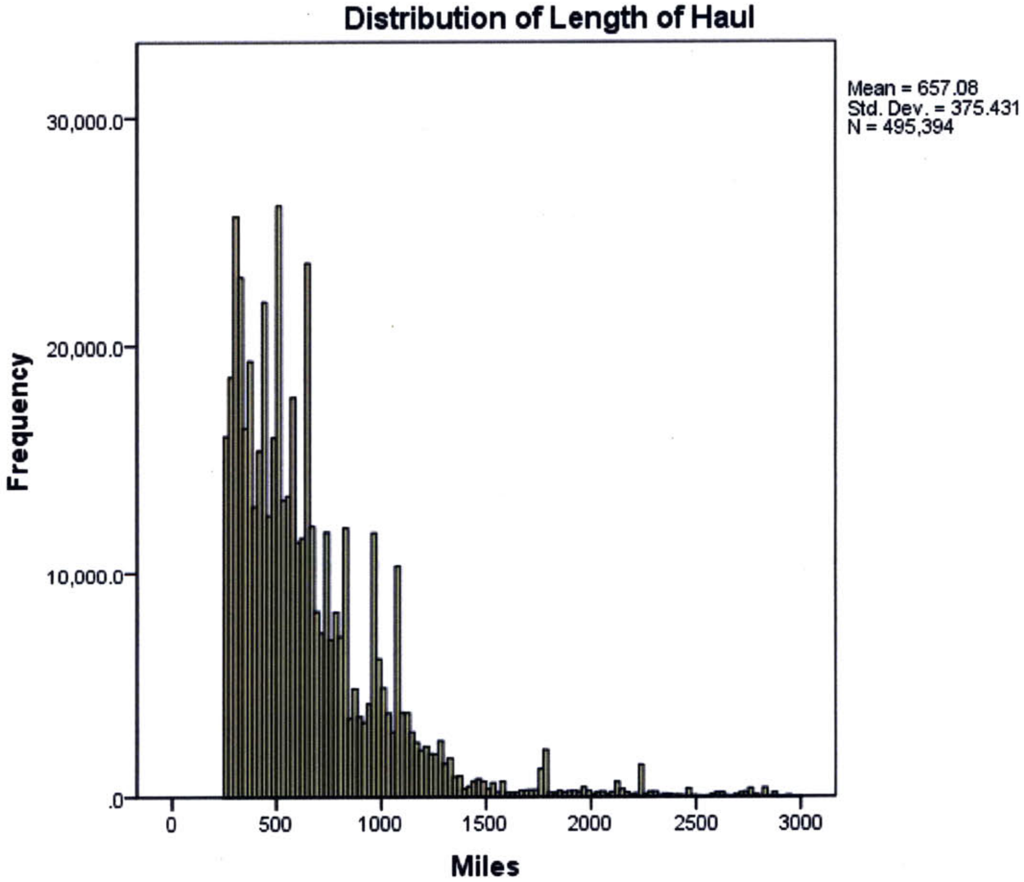


Figure 3: Distribution of Length of Haul

Figure 4 shows the distribution of cost per mile by load. The mean of $1.51 and standard deviation of $0.44 mean that 68% of tender rates have a cost per mile between $1.07 and $1.95, assuming a normal distribution.

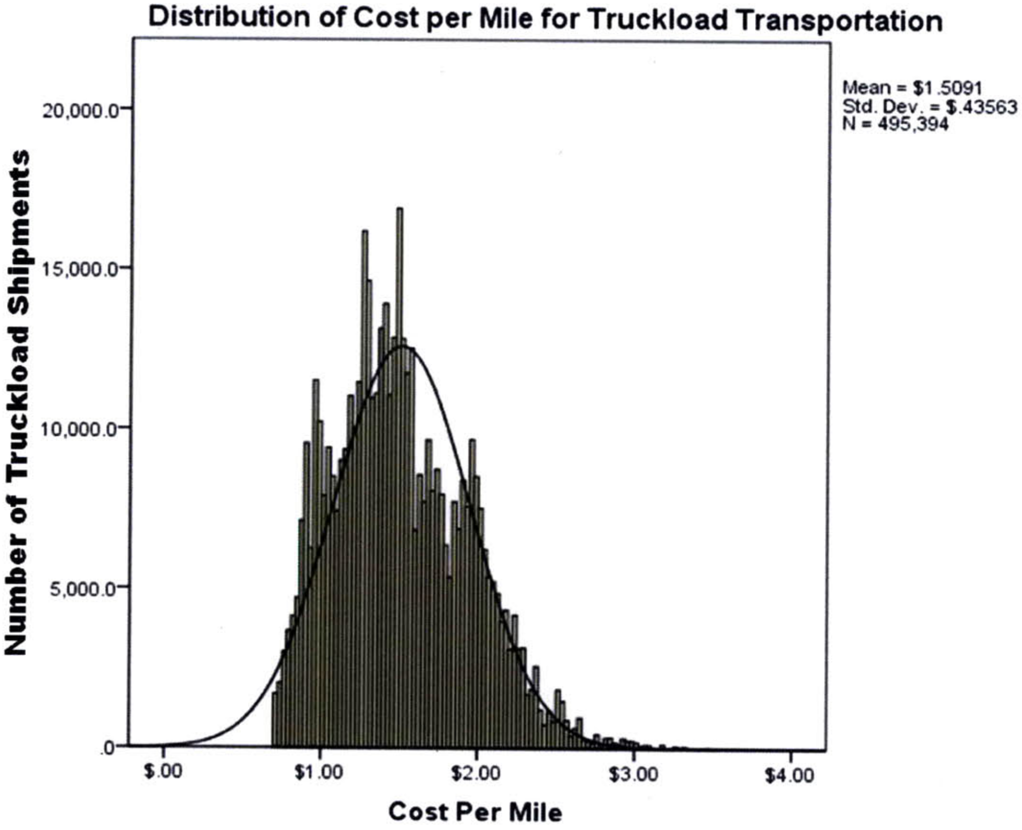


Figure 4: Distribution of Cost per Mile for Truckload Transportation

We sought to understand customer characteristics to better evaluate the model. We wanted to be conscious of the relative weight of an individual customer in the dataset. The number of loads per customer gave us a snapshot of how much each customer contributed to the data set. While we had eliminated customers with fewer than 2,000 loads per year in the years examined, Table 3 illustrates that the number of loads per customer covers a wide range. Customer 1 accounts for 40% of the total loads, with customer 2 accounting for 20% more. Although customers 1 and 2 account for the majority of the loads, the characteristics of these customers mirror the characteristics of the data set as a whole.

Customer Number	Total Loads	Avg. Length of Haul	Avg. Cost Per Mile	Spot Loads	% Spots
1	201,443	675	$ 1.61	3219	2%
2	104,983	593	$ 1.38	3117	3%
6	9,325	521	$ 1.34	1240	13%
7	52,728	704	$ 1.57	18643	35%
11	43,551	620	$ 1.47	3876	9%
16	35,797	634	$ 1.50	400	1%
17	4,647	1111	$ 1.36	280	6%
24	42,920	712	$ 1.36	721	2%
Total	**495,394**	**696**	**$ 1.45**	**31,496**	**6%**

Table 3: Customer Profiles

Figure 5 shows that while the customers differ in volume shipped there is consistency in how the average length of haul affects the cost per mile. The one outlier in Figure 5 is customer 17. Customer 17's average length of haul is significantly greater than the other customers without any significant change in the cost per mile.

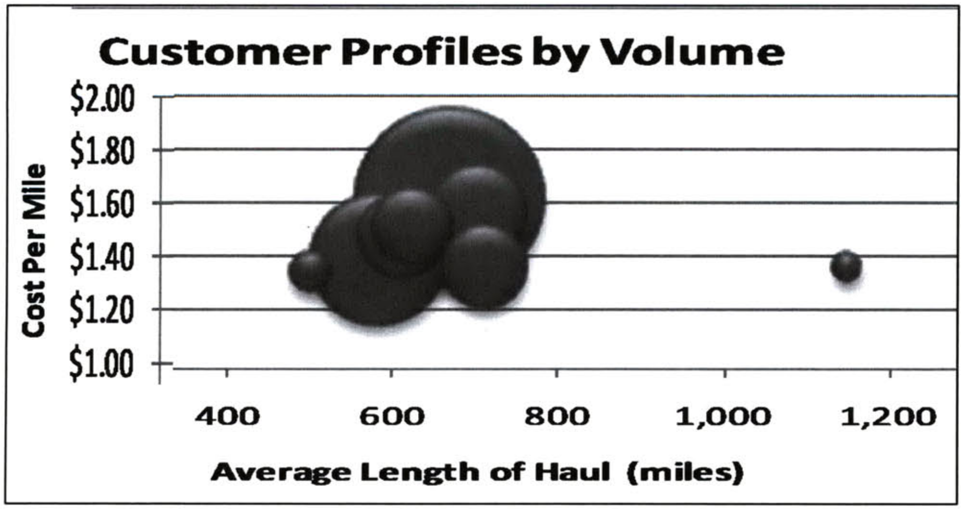


Figure 5: Customer Profiles by Volume

As seen in Table 3, most customers use the spot market sparingly. The notable exception is Customer 7, who uses the spot market for about 35% of its total loads. The total number of spot bids relative to the total number of loads shipped, 31,496 out of 495,394 loads, is substantial. The proportion of the total loads bid on the spot market suggests that it is necessary to consider contract type as a variable in the linear regression. Each lane was bid out as a specific type of origin or destination. Figure 6 shows the breakdown of bids by origin-destination combination. Table shows bids by origin type and destination type, with City to City clearly the most popular type of bid. Origins were bid as City in almost 90% of total bids. The next most popular bid type was spot bids. The overwhelming majority of bids use point origin or destination types, either 5Zip, City, or Warehouse, a specific address, which suggests that point is the default for carriers. Specifically, the bid types suggest, and CHR confirms, that customers typically have specific origin points in mind. The strong presence of spot bids suggests that the lanes that the shippers bid out are not collectively exhaustive, necessitating extensive spot bidding.

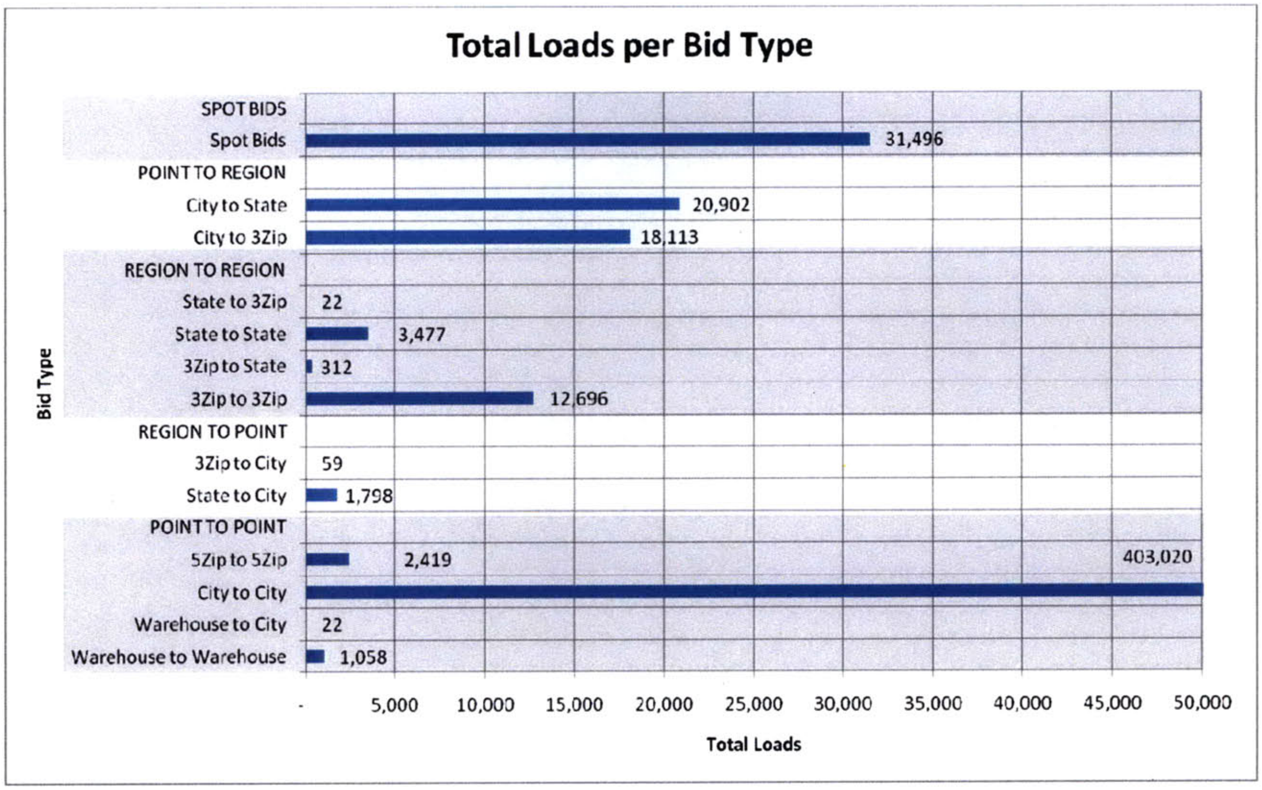


Figure 6: Total Loads by Bid Type

			Destination Quote Type						
				Point			Region		
			Spot	Warehouse	5Zip	City	3Zip	State	TOTAL
Origin Quote Type		Spot	31,496	-	-	-	-	-	**31,496**
	Point	Warehouse	-	1,058	-	22	-	-	**1,080**
		5Zip	-	-	2,419	-	-	-	**2,419**
		City	-	-	-	403,020	18,113	20,902	**442,035**
	Region	3Zip	-	-	-	59	12,696	312	**13,067**
		State	-	-	-	1,798	22	3,477	**5,297**
		TOTAL	**31,496**	**1,058**	**2,419**	**404,899**	**30,831**	**24,691**	**495,394**

Table: Loads by Origin and Destination Bid Type

5 Analysis of Data

This section explains the analysis of the data and the results. Using multi-variable ordinary least squares regression we tested over 50 different base models to show the effects of lane aggregation. This section discusses the various models tested and explains the rationale for eliminating certain versions in favor of others.

5.1 Regression Analysis

We used ordinary least-squares multiple linear regression in statistically testing all three models. A linear regression model assumes the following (Bertsimas and Freund, 2004):

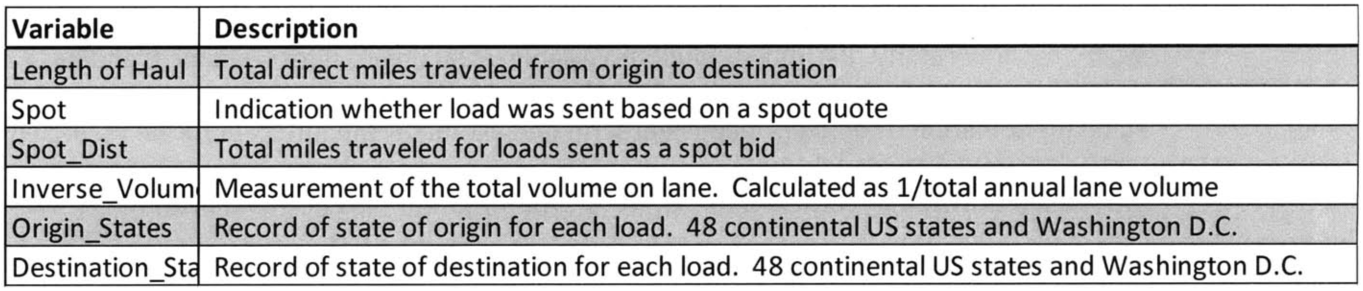
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon,$$

where ε is a normally distributed random variable. The mean, μ, of ε equals 0 and has a standard deviation of σ. β_0 is a constant term and estimates the intercept of the predicted model's path with the vertical axis. $\beta_1, \beta_2, \ldots, \beta_k$ represent the coefficients on the independent variables whose values determine the dependent variable Y. The independent variables are represented by $x_1, \ldots, x_k$. In our model the dependent variable, Y, is the line-haul rate in dollars per load. The independent variables include distance, geographic origin and destination, spot bid status, and the other methods variables captured aggregation influence.

The measurement used to evaluate how well the linear regression predicts the value of the dependent variable is the coefficient of determination. The coefficient of determination is represented by R^2. R^2 assesses the model's ability to predict dependent variable by evaluating the proportion of total variation of the values of the dependent variable that is accounted for by the regression equation of the independent variables (Bertsimas and Freund, 2004).

5.2 Building the Base Model

The purpose of our base model was to estimate line-haul contract rates for the U.S. for-hire trucking market. As stated earlier, line-haul rates are mostly driven by distance travelled and regional effects. Therefore, these factors were included as independent variables but others were also added so that the effects of aggregation are most clearly seen. The six independent variables used to make up the base model are listed in Table 5 with a brief explanation of each. Further explanation for inclusion in the base model is below.



Variable	Description
Length of Haul	Total direct miles traveled from origin to destination
Spot	Indication whether load was sent based on a spot quote
Spot_Dist	Total miles traveled for loads sent as a spot bid
Inverse_Volum	Measurement of the total volume on lane. Calculated as 1/total annual lane volume
Origin_States	Record of state of origin for each load. 48 continental US states and Washington D.C.
Destination_Sta	Record of state of destination for each load. 48 continental US states and Washington D.C.

Table 4: Description of Base Model Variables

5.2.1 Length of Haul

Length of haul is the factor with the single largest impact on price for long haul, truckload shipments. Length of haul explained 72% of the variability in the cost per load in all of the regressions run. This result is consistent with long term industry findings about the importance of length of haul in truckload trucking. The importance of length of haul provides a strong and consistent foundation for the model, because the majority of cost is based on this one factor. As explained earlier, trucking operates on essentially a cost-plus pricing model based on length of haul and geography.

In the model, length of haul is represented by Miles, which shows the cost per mile that must be added to the constant. Cost per mile was about $1.15 across all runs.

5.2.2 Regional Effects

Regional effects explain 17% of variability ($R^{2)}$ value of the dependent variable. In the model, regional effects were evaluated using 98 variables, one for each origin and destination zone. Assessing regional effects by state allowed us to define each region using a specific, easily identifiable area, while still being broad enough to show general trends.

The regional effects are a measurement of economies of scope. As described earlier, economies of scope have a significant impact on line-haul pricing. As shown in Figure 7 and Figure 8, shippers pay a premium based on origin and destination state. This explains 17% of the line-haul cost variance, as described earlier. As shown in the Figure 7, for example, a carrier would pay a higher premium to ship out of Minnesota than out of Florida. As destinations, there is a lower premium to ship into Minnesota than Florida. The findings here are consistent with conventional wisdom about shipping out of certain geographies. There is a higher volume of shipments originating in Minnesota than bound for Minnesota, and the opposite is true for Florida. So, it is not surprising that the premium would be higher to ship into Florida than Minnesota. Figure 8 shows the premium by destination state.

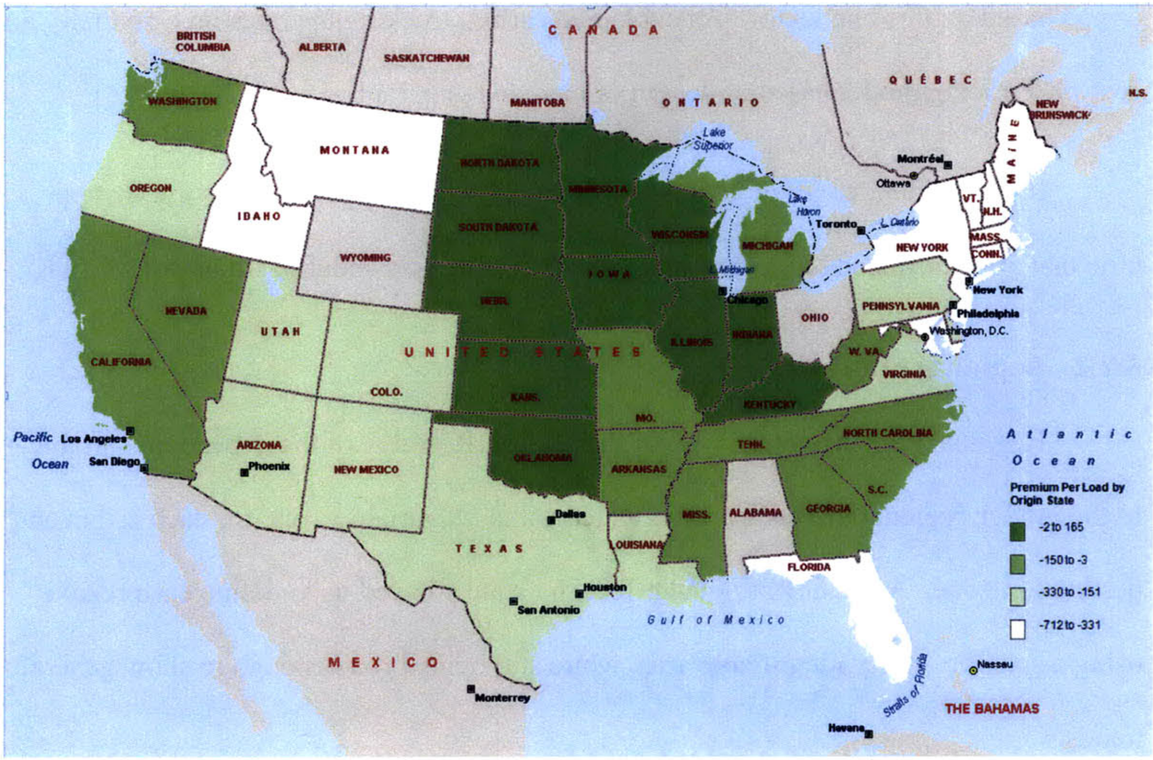


Figure 7: Premium per Load by Origin State

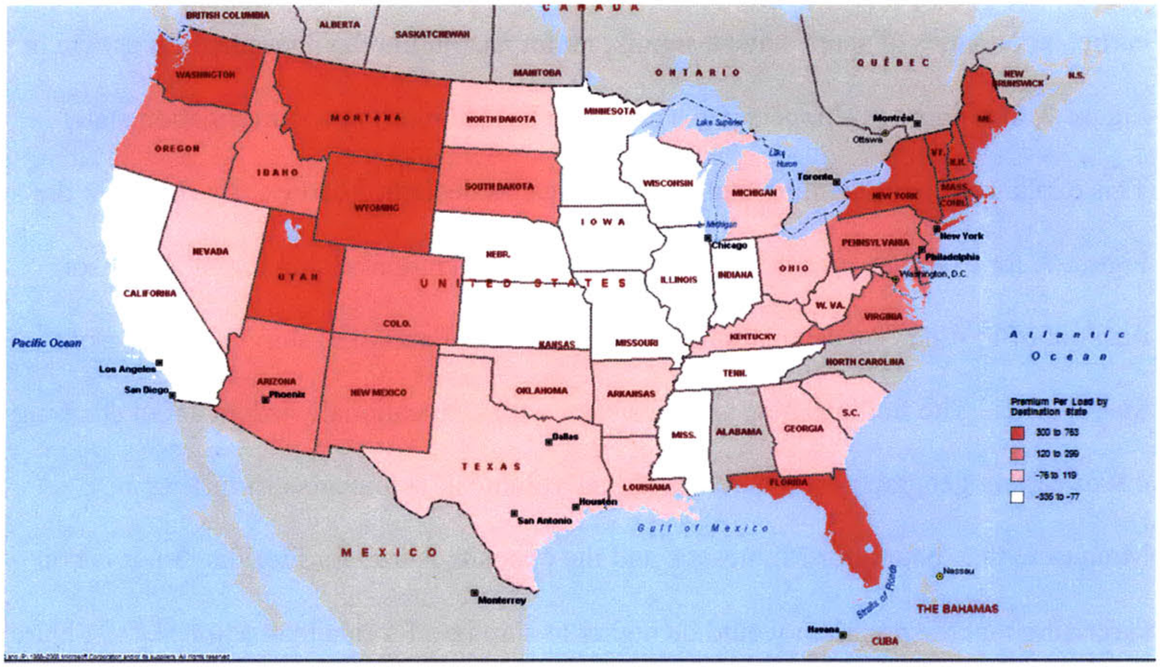


Figure 8: Premium per Load by Destination State

5.2.3 Equations Tested

Using the key variables described above, we used Length of Haul (Miles) and Regional Effects (the origin and destination zones) to explain 90% of the variability in the data set. After adding these two sets of variables we set out to add additional variables that could explain different facets of the lane aggregation question. Because the R^2 value was already at .904, any increases in R^2 were not as important as the value and statistical significance of the coefficient variables.

5.2.4 The Base Model

The first set of variables added is listed in Table 5 below. Spot, Spot_Dist, and Inverse_Volume were all added successively to the initial model of Miles and Origin/Destination States. All of the coefficients for each origin and destination state are not listed in the exhibits in the main body of the thesis. Because this thesis is concerned with the question of the impact of aggregation, listing the coefficients for each of the states does not help to support that goal. These can be found in Appendix 1.

The coefficients for the variables listed in Table 5 show the individual effects for each of the variables. For example, looking at Run 4, the effect of each mile driven is to add $1.15 to the cost of the load. In addition, if a lane only has one shipment on it over the two-year period that the dataset represents then there will be a premium on that specific load of $60.43. If a contracted load travels 1,000 miles and has a total of one shipment a year on the lane then the cost of the load will be $1,498.69 (without including the effects of origin and destination states).

RUN Number	1	2	3	4
R-Squared	91%	91%	91%	91%
(Constant)	$ 303.01	$ 307.97	$ 317.91	$ 319.47
Miles	$ 1.17	$ 1.17	$ 1.15	$ 1.15
Spot		$ 111.08	$ 7.41	$ 3.09
Spot_Dist			$ 0.13	$ 0.13
Inverse_Volume				$ 60.43
Origin States Included	X	X	X	X
Destination States Included	X	X	X	X

Table 5: Base Model Development

Table 6 shows the detail behind this specific calculation. The power of the regression's coefficients is that each variable's effect can be specifically measured as in this example. For example, these two sample shipments demonstrate how each of the individual variable coefficients affects the predicted price for the line-haul cost. This specific example also demonstrates the impact of relative location on total cost. Because Florida does not have a great deal of freight coming out of it, carriers are reluctant to go enter because they know how difficult it will be to find a backhaul. Therefore, there is a premium on loads going into Florida.

	Chicago to Miami		Miami to Chicago	
	Values	Cost	Values	Cost
Constant		$ 319.47		$ 319.47
Miles	1375	$ 1,581.25	1375	$ 1,581.25
Spot	0	$ -	0	$ -
Spot_Dist	0	$ -	0	$ -
Inverse_Volume	2	$ 30.21	2	$ 30.21
Origin_State		$ 4.20		$ (389.20)
Destination_State		$ 259.80		$ (226.50)
Total		**$ 2,194.93**		**$ 1,315.23**

Table 6: Example of Base Model Prediction

5.3 Testing Economies of Aggregation

In this section we discuss first how to represent economies of aggregation and then quantify the impact.

5.3.1 Representing Economies of Aggregation

In order to build a model that was detailed enough to evaluate aggregated bidding effects, we separated each of the variable's effects. We used three general approaches to understand aggregation's impact on truckload bidding. These three approaches tested the effects of region size, lane volume, and empty miles distance. We first tested the impact of region size on tender rate. The original data included the bid type in specific detail showing whether the origin and destination points were bid as either warehouse, 5 digit zip code, city, 3 digit zip code, or state. The relative size of the geographic region is generally summarized in the order just listed with warehouse being the smallest geographic point and state being the largest. Because the data included this level of detail it made sense to test each type's effect on truckload pricing.

We also represented the effects of region size with the generic categories "point" and "region." This was done because the differences between some of the regions were not significant enough to warrant deep separation. The categories were thus represented by point-to-point, point to region, region to region, and region to region. Our contact at CHR suggest that bid types warehouse, 5 digit zip code, and city all behave like a single point and could therefore be defined as such. The reasoning behind this distinction was that these three areas are small enough that experienced truckload carriers will have a very good idea of where each load will be picking up and delivering and will therefore bid out the lane as a specific point. For 3 digit zip codes and states introduced significant ambiguity for the carrier. Size of a region indicated that carriers would be forced to look at the possibility of driving empty miles either at the origin or destination to complete the load.

The second approach to understanding aggregation's effect on bidding is representing the volume of loads on each lane. We defined a lane as a 5 digit zip code to a 5 digit zip code. We used the "Inverse_Volume" variable to represent the effect of adding an additional load to each lane. The inverse volume variable measures the incremental effects of volume on rates.

The third approach was to represent the number of empty miles on each load. Running an empty truck is often referred to as "deadheading" and can happen either at the origin or destination of a load. We had to determine the size of each origin and destination region, by 5 digit zip code, 3 digit zip code, and state, and then determine how far a carrier would likely have to drive empty.

We gathered data on the square mileage of each five digit zip code from the United States Census Bureau. We chose to use information from the Census Bureau because it provided a complete record of square mileage for five digit zip codes for the whole United States. This information was not readily available from any other source, much less one so reliable. There was one major drawback to the Census data: the square mileage includes only land area, excluding water area in a given location. However, the measuring methodology was consistent throughout the country and provided an exhaustive record of the nation's land area. We used this data to find the area of each 3Zip in our data, as well. We also used it to find the area of each state in our data. While for the square mileage of each state is readily available from the U.S. Geological survey and other reliable sources, we chose to be consistent in our area calculations. There was some discrepancy between official square mileage at the state level and the square mileage we found using the Census data; in each case, the Census data provided a smaller area. However, the differences were not significant, and consistency provided a more compelling option than precision. As stated above, all origins and destinations bid as Warehouse, City, or 5Zip were assigned an area of zero square miles.

To capture this effect we used the centroid technique. The idea of a centroid is to measure the average distance that a carrier would have to travel either at the origin or destination. We estimated each load's centroid by the area of its 5 digit zip code, 3 digit zip code, or state based on how the load was bid out. The intent was to capture the carriers' reaction to being asked to bid on a lane knowing different levels of information at both the origin and the destination. Empty miles were estimated for the origin, destination, and the complete lane.

To calculate the average distance from the centroid, we assumed that a carrier would have to move an average of half the distance from the centroid to the edge of the zone. That is based on the assumption that the greatest distance that carrier would move within a zone would be from the edge to the centroid, and the least would be no distance. We assumed that the zones took a uniform shape, either square or circular, because those two shapes made the most practical sense when estimating for driving distances. We found it necessary to assume uniform shape for the zones as it would be unfeasible to calculate the average distance from the centroid for the zones in their actual shapes. There are two reasons for this. One, the shapes of the zones are so irregular that it would require excessive time to calculate this value for each zone. Two, information about the shape of each zone is not readily available. Using the following calculations allowed us to follow a quick and uniform method for calculating average distance from the centroid in miles. We tested the average distance from the centroid using each shape separately in the model.

For the zones as squares, we calculated the average distance from the centroid as $d = \sqrt{\frac{A}{8}}$. We derived the formula as follows. We identified distance from the centroid, *d*, as half the length of the side of the square that is half the area of the zone. So, length of side equals *2d*. We are assuming that half the area of the zone is also square. The area of half the zone equals$(2d)^2 = 4d^2$. This represents half the area of the zone, $\frac{A}{2}$. So, $\frac{A}{2} = 4d^2$. We can then solve for *d*. $\frac{A}{8} = d^2$. So, $d = \sqrt{\frac{A}{8}}$.

We calculated the average distance from the centroid for the zones as circles using the same methodology. We again assumed that average distance from the centroid to be based on the distance from the center to the edge of a circle with half the area of a zone. Using the distance, *d*, for the radius, the area of the circle with radius *d* is $A = \pi d^2$. Because this represents half the area of the zone, $A_{zone} = 2\pi d^2$. $\frac{A}{2\pi} = d^2$. So, solving for *d*, we find $d = \sqrt{\frac{A}{2\pi}}$.

There are drawbacks to these assumptions. The calculation for area of a square region results in a longer average distance from the centroid than for a circular region based on the basic geometry of the shapes. As with square areas of the regions, which we know are smaller than the regions actually are, the values generated by these calculations are imprecise, but calculated consistently.

We further defined the effect of empty miles by creating a ratio of total miles to estimated empty miles. Variables were created for origin, destination, and the complete lane. These variables are intended to capture the effect of empty miles and simultaneously account for the effects of the length of the haul.

5.3.2 Quantifying the Impact of Economies of Aggregation

Calculating the effects of lane aggregation required several different approaches. As stated earlier, these efforts can be broadly categorized as the effects of region size, lane volume, and empty miles. The additional variables are defined in Table 7 below:

Variables	Description
Point_Region	General categorization of origin and destination as either a point or a region.
Region_Point	
Region_Region	
OriginEmptyMilesSquare	Estimate of empty miles that TL must drive at the origin, destination, and total lane for square region.
DestEmptyMilesSquare	
LaneEmptyMilesSquare	
OrigLHRatioSquare	Ratio of line-haul miles to total empty miles for square regions.
DestLHRatioSquare	
LaneLHRatioSquare	
OriginEmptyMilesCircle	Estimate of empty miles that TL must drive at the origin, destination, and total lane for circle region.
DestEmptyMilesCircle	
LaneEmptyMilesCircle	
OrigLHRatioCircle	Ration of line-haul miles to total empty miles for circle regions.
DestLHRatioCircle	
LaneLHRatioCircle	

Table 7: Explanation of Variables

Out of the approximately 50 regressions we ran, we ultimately chose the final model from among 10 runs. We will discuss each run individually to explain why we did or did not choose it for the final model. A full table of the runs can be seen in Appendix 2.

As previously stated, the data for this study ranged from 2008-2009. Because of general economic conditions in the U.S., both of those years were "down" years for trucking. A down year is one in which there are more trucks available then freight needing to be moved. As a result of simple supply and demand economics, this should result in lower truckload pricing. During a down market there is less volume going to the spot market because contract carriers are not rejecting as much business. This is consistent with our data, which shows that spot shipments dropped by almost 80% from 2008 to 2009, from roughly 26,000 in 2008, to just under 6,000 in 2009.

The range of coefficients for Spot and Spot_Dist range from \$3.25 to \$13.71 and \$0.13 to \$0.14 respectively. Therefore, during the 2008-9 down market, spot rates cost more than contract rates.

We selected Run 7 as the best model for this research. Run 7 included a very reasonable value for the Inverse_Volume variable as well as detailed explanatory coefficients for both the origin and destination empty miles variables. Run 7 was chosen over Run 9, which shows the effects of empty miles over the lane, because it isolates the effects of empty miles at the origin and the destination even though the same overall result is evident. Run 7 is essentially the same as Run 8 and the two are effectively interchangeable. The only difference between the two is the assumed shape of regions. Run 7 assumes that the regions are square, while Run 8 assumes that they are circular.

The same arguments and conclusions that will be drawn from Run 7 can also be drawn from Run 8. Run 7 will be referred to as "the model" going forward in this paper.

RUN Number	**7**
R-Squared	91%
(Constant)	$322.31
Miles	$ 1.14
Spot	$ 9.68
Spot_Dist	$ 0.14
Inverse_Volume	$ 26.78
OriginEmptyMilesSquare	$ 0.91
DestEmptyMilesSquare	$ 0.50
Origin States Included	X
Destination States Included	X

Figure 9: Run 7, the Model

The model provides enough information to be able to describe the effects of aggregating low-volume lanes and gives specific numbers on both the origin and destination of the load to help define the optimal area size. Simply defining larger areas as "regions" and smaller areas as "points" does not give enough information for companies to be able to optimize region or point geographic sizes. A premium tolerance is the level of premium that a company is willing to pay per load on a specific model variable to provide a benefit or flexibility in another way to the company. For example, a shipper may decide to make aggregation levels large in order to reduce cost of managing many smaller lanes. The results of the model give this company the information needed to calculate the estimated amount that making such a decision would cost.

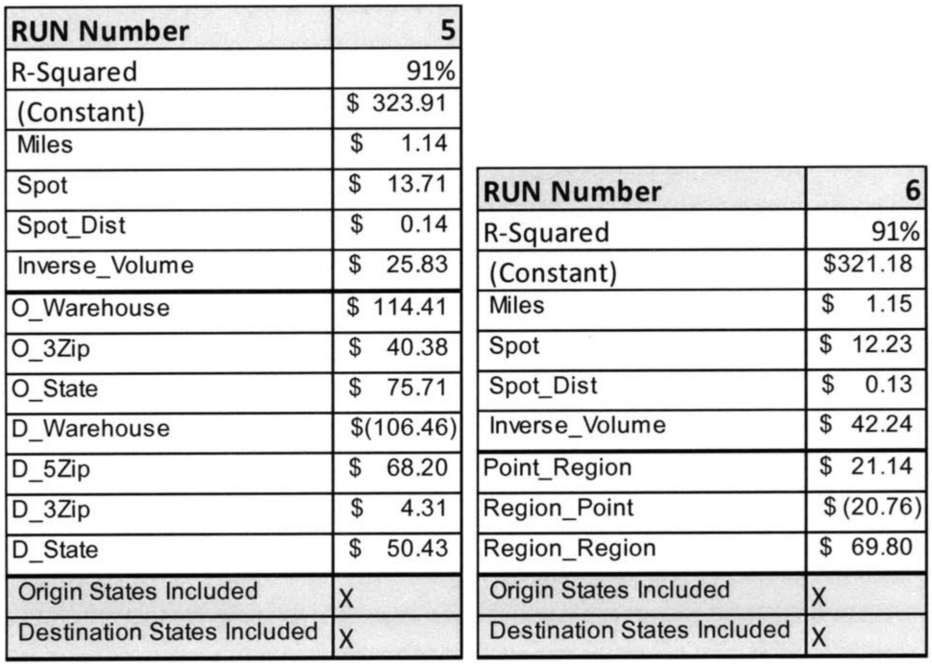
Run 7 also held consistent results with the base model. For Miles, Spot, Spot_Dist, and Inverse_Volume the results were consistent with the majority of the other runs. This cannot be said about some of the other runs. For example, the runs that switched the empty miles variables to ratio, whether square areas or circle, experienced dramatic shifts in the Constant, Spot, and Inverse_Volume variables. Run 7 did not have similar problems in the base case.

5.3.3 Other Runs Considered

There were nine other runs that provided valid, significant results. Having chosen Run 7, this section will examine why we did not choose the other runs. As explained above, the results from runs for which the only difference was the region shape (Square or Circle). This was the case for Runs 7 and 8, 9 and 10, 11 and 12, and 13 and 14.

5.3.3.1 Regional Size Effects

Run 5 shows the effects of each specific bid type on the model. City as Origin and Destination do not appear in the regression because they are the base case for the model. We rejected this run as we had determined that the distinctions between individual point and region bid types were artificial. In Run 6, the individual bid types of warehouse, 5-digit zip code, 3-digit zip code, and state are accounted for. All are self-explanatory except for possibly warehouse. In this dataset, a warehouse was defined as a specific street address and simply categorized as a "warehouse" because so many of the examples in this category were street addresses of warehouses. However, we found that the results were not specific enough. Since 3 digit zip codes are treated the same way as states, this means that a 3 digit zip code in the Dallas area is treated the same way in the model as the entire state of Texas. Also, Run 6 shows that modeling the regions so broadly influences the inverse volume effect. In this run, Point to Point serves as the base case for the model. For most other runs tested, the inverse volume effect is roughly $25.00 for the lane. In Run 6, however, the effect is $42.24. This deviation suggests that defining the lanes so broadly distorts the other variables in the model.



RUN Number	5
R-Squared	91%
(Constant)	$ 323.91
Miles	$ 1.14
Spot	$ 13.71
Spot_Dist	$ 0.14
Inverse_Volume	$ 25.83
O_Warehouse	$ 114.41
O_3Zip	$ 40.38
O_State	$ 75.71
D_Warehouse	$(106.46)
D_5Zip	$ 68.20
D_3Zip	$ 4.31
D_State	$ 50.43
Origin States Included	X
Destination States Included	X

RUN Number	6
R-Squared	91%
(Constant)	$321.18
Miles	$ 1.15
Spot	$ 12.23
Spot_Dist	$ 0.13
Inverse_Volume	$ 42.24
Point_Region	$ 21.14
Region_Point	$ (20.76)
Region_Region	$ 69.80
Origin States Included	X
Destination States Included	X

Figure 10: Cost Effect of Bid Type on Line-Haul Cost

As explained earlier, we first represented regional size effects with the origin/destination bid types. As explained above, region size was also modeled more generically with categories representing either a point or a region for both origin and destination. The results shown in Run 5 are interesting but not complete enough to explain the total effects of lane aggregation.

The results for the different origin/destination bid types assume the "city to city" origin destination pairing as the base case. We designed the model this way because this bid type represents the majority of bids in the dataset. Therefore, the coefficients for each of the variables are relative to the city to city bid type. Also, the regression program automatically dropped the 5-digit zip code origin loads, because all loads originating with a 5-digit zip code were destined for a 5-digit zip code. Also, all loads destined for 5-digit zip codes originated in 5-digit zip codes. Hence, having both variables in the model was redundant since one variable inherently represented the effect of both.

There is insufficient data from each origin/destination type in this dataset to be substantial enough for us to extrapolate results to the entire for-hire TL industry. Many origin/destination type pairings do not have any data points. To account for this lack of data points, we broadened the definition of the origin/destination type to simply be "point" or "region."

In the point and region comparison in Run 6 the "point-to-point" pairing is used as the base case and the other pairings coefficients are relative to point-to-point. This general classification does more to explain the effects of aggregation on low-volume lanes. For example, the results show a $21.52 cost for load going "point-to-region" and a $22.77 savings for loads going "region-to-point." This effect may be caused by carriers' desire to get close to a certain place when delivering a load. This could be because the destination is close to other available loads or it could simply be the home city of the driver. In any case, the results show that carriers could be willing to offer companies a $22.77 savings per load if the shippers broadly define the origin point. Regarding lane aggregation, this could mean that shippers with multiple lanes originating from multiple locations in a general area moving to a specific destination may be able to save $22.77 per load by aggregating those loads at the origin. However, the results also imply that companies will lose $21.52 by aggregating at the destination. Therefore, a more complete independent variable or variable set must be used to clarify these implications.

From the "region to region" variable it is clear that defining the areas too broadly has costs. But as with the other two variables evaluated above, it is not specific enough for companies to use for strategy development. These variables could be best used as broad guidelines that suggest aggregating may have benefits, but they do not supply the detail necessary to implement aggregation strategy changes.

5.3.3.2 Volume Effects

The Inverse_Volume variable has already been explained above. The results of the variable range from $24.88 to $59.41 depending on which regression run the independent variable is included in. Therefore, this variable must be evaluated based upon how its cost effects complement and complete the picture that the other variables included in the model create. In relation to the two variable sets for region size, the volume effect goes from $25.83 to $42.24. This suggests that the premiums assigned to the origin and destination are absorbing some of the effects of volume since the volume premium in that particular run is lower. It may also suggest that the general categorizations of "point" and "region" may be too broad. If the volume variable changes by more than 50% when defining region size more broadly, then the model is in a weaker position to fully explain how increased volume impacts line-haul costs on low-volume lanes.

	N	Minimum	Mean	Maximum	Std. Deviation
OriginEmptyMilesSquare	499,272	0	1.40	181	10.1
DestEmptyMilesSquare	499,272	0	5.34	181	20.3
LaneEmptyMilesSquare	499,272	0	6.74	362	25.2

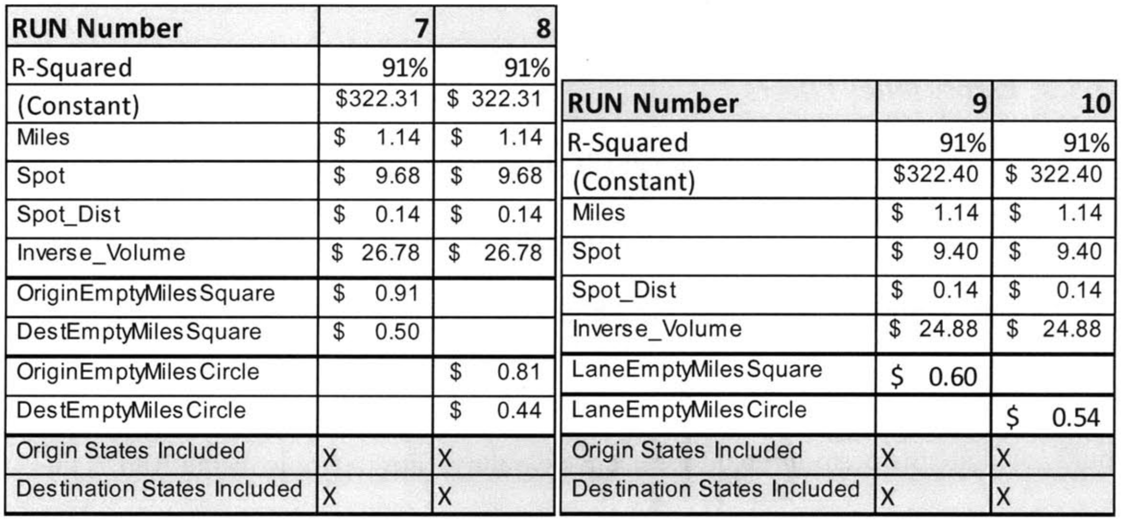
Table 8: Descriptive Statistics for Empty Mile for Square Regions

The last point with this set of variables is the influence on the model itself. As shown in Table 8, the distribution of these variables makes the highest possible premium \$59.84 (\$0.061 x 981).

5.3.3.3 Empty Miles Effects

As stated above, Run 8 gave effectively identical results to Run 7. Runs 9 and 10 gave results that were not significantly different from Runs 7 and 8, but they show the effect of empty miles driven across the whole lane. That is, if the empty miles driven at the origin were 5 and the empty miles driven at the destination were 100, Runs 7 and 8 would calculate the premium paid for each separately. For example, using Run 7, the premium would be: $\$0.91 * 5 + \$0.50 * 100 = \$93.50$. However, using Run 9, which assigns the same price premium to empty miles at the origin and destination, the premium would be calculated as $\$0.60 * 105 = \$63.00$. Comparing Runs 8 and 10 would yield the same type of result. The greater specificity of Runs 7 and 8 give a more accurate result, which is why Run 7 was chosen rather than Run 9 or 10.

Empty miles are a key driver in line-haul rate for TL carriers. Also outlined above is the method used to model empty miles within this study. The results for the coefficients for both a square area and a circular area tell the same story. Both show almost a 50% increase in premium for origin empty miles over destination empty miles. The coefficients are also virtually the same given the difference in area size between a square area and a circular area (\$0.90 and \$0.80 for origin empty miles for square and circle respectively and \$0.50 and \$0.44 for destination empty miles for square and circle respectively). The results are also practically the same for the combined lane empty miles for a square area and circle area (\$0.60 and \$0.53 respectively). These regression results also produce similar results on the volume coefficient variable (ranging between $24.88 and $26.78).



RUN Number	7	8
R-Squared	91%	91%
(Constant)	$322.31	$ 322.31
Miles	$ 1.14	$ 1.14
Spot	$ 9.68	$ 9.68
Spot_Dist	$ 0.14	$ 0.14
Inverse_Volume	$ 26.78	$ 26.78
OriginEmptyMilesSquare	$ 0.91	
DestEmptyMilesSquare	$ 0.50	
OriginEmptyMilesCircle		$ 0.81
DestEmptyMilesCircle		$ 0.44
Origin States Included	X	X
Destination States Included	X	X

RUN Number	9	10
R-Squared	91%	91%
(Constant)	$322.40	$ 322.40
Miles	$ 1.14	$ 1.14
Spot	$ 9.40	$ 9.40
Spot_Dist	$ 0.14	$ 0.14
Inverse_Volume	$ 24.88	$ 24.88
LaneEmptyMilesSquare	$ 0.60	
LaneEmptyMilesCircle		$ 0.54
Origin States Included	X	X
Destination States Included	X	X

Figure 11: Cost Effect of Empty Miles Traveled at Origin and Destination

Runs 11 and 12 show the cost effect per mile based on the ratio between the distance traveled at the origin and destination and the line-haul distance, using the variables OrigLHRatioSquare, OrigLHRatioCircle, etc. Once again, the results for treating the zones as square or circular are interchangeable. The results of these two runs do not support practical knowledge of the relationship between line-haul distance and distance traveled at the origin or destination. Intuitively, the larger the line-haul distance compared to distance traveled empty, the smaller the price premium for traveling empty miles should be. The model does not accurately capture this effect. Runs 11 and 12 also significantly distort the other variables in the model. The constant, which for other runs was around $320, for these two runs dropped to $56. We rejected the results of Run 13 and 14 for the same reason.

RUN Number	11	12
R-Squared	91%	91%
(Constant)	$ 56.42	$ 56.42
Miles	$ 1.15	$ 1.15
Spot	$ 3.99	$ 3.99
Spot_Dist	$ 0.13	$ 0.13
Inverse_Volume	$ 58.50	$ 58.50
OrigLHRatioSquare	$ 0.17	
DestLHRatioSquare	$ 0.11	
OrigLHRatioCircle		$ 0.20
DestLHRatioCircle		$ 0.13
Origin States Included	X	X
Destination States Included	X	X

RUN Number	13	14
R-Squared	91%	91%
(Constant)	$319.63	$ 319.63
Miles	$ 1.15	$ 1.15
Spot	$ 3.25	$ 3.25
Spot_Dist	$ 0.13	$ 0.13
Inverse_Volume	$ 59.41	$ 59.41
LaneLHRatioSquare	$ 0.06	
LaneLHRatioCircle		$ 0.07
Origin States Included	X	X
Destination States Included	X	X

Figure 12: Cost Effect of Line-Haul Ratio

5.4 Regression by Customer

We tested the model for each customer to see if it yielded similar results on the customer level. We used the same variables as Run 7 but ran it for each customer. The results for this exercise were surprising. For each customer, the base model explained most of the line-haul cost variance, from 91% to 97%. Customers 1 and 2, which represent over 50% of the total loads, exert a strong influence on the final model. All the variables were significant for these two customers. Their Constant values were by far the highest of the set, well above that for the dataset as a whole. The null values in Origin and Destination Empty Miles reflect the customers' bidding strategies; their bids were all made as points. For Customer 6, the Destination Empty Miles was insignificant as were the Origin Empty Miles for Customer 24. This reflects that these two customers bid a few lanes with region destinations and origins, but that there was not enough volume of these bids to be significant to the model. Overall, the conclusion that can be drawn from the results of the regression by customer is that it is impossible to make generalizations about the market based on a single customer's data. The model's robustness is tied with the depth and breadth of data from a variety of customers. Based on the full dataset, however, the model can be applied to each customer.

The model we have developed has yielded results that allow us to quantify the correlation between lane volume and line-haul rates. We tested several ways to assess the effect of volume, including by specific and general bid types, but concluded that the Inverse_Volume gave the best result because it allowed us to capture the impact that each additional move on a lane had on cost. In testing several methods to quantify cost of deadhead at the origin and destination, we were able to capture the cost of each empty mile driven at the origin and destination. Because aggregation can be done at the origin, destination, or both, capturing these cost effects individually has greater value than capturing this cost effect for the lane as a whole, and can assist a shipper in evaluating where and when aggregation would be most appropriate.

6 Conclusion

Aggregating lanes can lower truckload costs. However, the size of the aggregation region at the origin and destination will determine whether it would be cost-effective to aggregate. Whether aggregation will yield cost savings is determined by the alternative to aggregation. If a shipper is deciding whether to bid a lane as point-to-point or aggregate it, then one set of criteria should be used. The model can also be used to test whether aggregation would yield cost savings by eliminating or reducing spot bids. This second use of the model relies on different variables within the same model. While the company itself will understand its specific needs and internal obstacles to aggregating certain lanes, aggregation can either positively or negatively influence prices. Companies will also have to decide internally how much time savings there is in aggregating versus having point-to-point lanes across a region. The regression model created in this study shows the estimated savings by extra load per lane or the cost by empty mile per location.

6.1 Key Variables and Variable Application

Ultimately, there are five key variables in our model that describe the relationship between the relationship between aggregation and empty miles: Inverse_Volume, OriginEmptyMilesSquare, DestEmptyMilesSquare, Spot and Spot_Dist. Our model shows that Inverse_Volume's coefficient is $26.78. Therefore, the dollar amount that lane volume adds to a given load is calculated by dividing $26.78 by the lane volume for that load. Lanes with only one load over the two-year period would have a $26.78 premium and lanes with only two loads, or an average of one load per year, have a $13.39 premium. If it is possible to aggregate these single- and double-load lanes to 25-load lanes over the two years, or 12-13 loads per year, the premium would only be $1.07 per load.

We understand that these results are only known after the fact, which is a limitation of the model. The model is not generic, and how many loads will move on a given lane per year is information that is not known to the shipper or carrier ex ante. To account for network externalities, other users must rerun the model using their own data.

Empty miles add an extra $0.90 coefficient at the origin and $0.50 coefficient at the destination. The coefficients for both of these variables are calculated assuming a square region using $\sqrt{(\frac{A}{8})}$. If a company decides to size a region at 100 square miles at the origin then there would be a $3.54 premium on each of the loads coming out of that lane as shown: $\sqrt{(\frac{100}{8})}$ = $3.54.

The relationship between premium on cost per load and the average distance from the centroid can be used together to find the region size at which costs for a point-to-point lane are equal to the cost to bid the lane as part of a zone or region. As shown above, the price premium for a single shipment per annum on a contracted lane is $13.39. To find the area at which the price premium based on distance driven at the origin is equal to $13.39, we can set $13.39 equal to the formula for average distance from a centroid, using square as the region shape multiplied by the price premium per mile for empty distance driven at the origin, as below:

$$\$13.39 = \$0.50\, x\sqrt{(\frac{A}{8})} \Rightarrow \$26.78 = \sqrt{(\frac{A}{8})} \Rightarrow 717 = \left(\frac{A}{8}\right) \Rightarrow A = 5{,}737\ square\ miles$$

So, if lanes that have a single shipment per year can be aggregated into a region that is smaller than about 5,740 square miles, or about 75 by 75 miles, then it will be more cost effective to aggregate with another lane or lanes if there are any in proximity to the lane evaluated. The more loads that move on a point-to-point lane per year, the smaller the area it would be cost effective to aggregate into. For example, if a lane had four loads per year, then the premium per load would \$26.78/4, or \$6.70. Using the same calculation as above, it would make sense to aggregate this lane only if other lanes could be found within a zone of 1,434 square miles.

$$\$6.70 = \$0.50\, x \sqrt{\left(\frac{A}{8}\right)} => \$13.39 = \sqrt{\left(\frac{A}{8}\right)} => 179 = \left(\frac{A}{8}\right) => A = 1{,}434\ square\ miles$$

1,434 square miles is an area of about 38 by 38 miles. So, the more loads per lane, the less incentive a shipper has to aggregate to realize any cost savings. Performing this calculation for the destination would yield even smaller region sizes, because the price premium at the destination is \$0.91, almost twice the premium of \$0.50 at the origin.

A specific example is the state of California. If a load is bid from a specific city to the state of California then our model gives the tools necessary to estimate the premium on each of the loads that are sent from that point to the general region of California. California is 156,000 square miles so the destination premium would be \$70.00. Therefore, aggregating at this large an area would have costly consequences since the most the low-volume premium will ever be is \$13.39. So the cost of aggregating at such a large level would always exceed the savings from adding more volume to the lanes. Companies looking to optimize savings would be better off choosing a smaller sized region or else the cost of the empty miles will outweigh the benefits. The economies of scale created by volume on a point to point lane dwarf any economies of aggregation created by larger regions. The result of our model is very sensitive to annual volume across a lane. As shown here, it is also sensitive to region size. Common sense suggests that aggregating to a region the size of California would be impractical, both due to its sheer size and the diversity of regions within that larger region. The Los Angeles area would have different market conditions from the Salinas Valley and the northern part of the state. Shippers and carriers would be aware of this and would in reality be unlikely to propose the state of California as an origin or destination region.

6.2 Impact and Implications of the Base Model

The variables that make up the base model support generally accepted characteristics of the for-hire, TL market. The main drivers of line-haul price are distance and location. These two factors alone explain 91% of the variance in line-haul cost.

The constant of $322 is also consistent with general findings in the market. TL rates are generally bound on the lower end of the distribution by a minimum cost. The minimum cost in this model is calculated by adding the constant to the product of the minimum possible miles driven and the coefficient for the variable miles. Since we eliminated all loads below 250 miles, this results in a minimum cost of $607.31 ($322.31 + 250*$1.14). As shown in Table 10 below, there are loads that cost below the $607.31 minimum that our model predicts. Because this study is not focused on how minimum costs are charged or how backhauls are priced we have not chosen to focus on the full spectrum of implications in these findings.

	N	Minimum	Maximum	Mean	Std. Deviation
Tender_Rate	495,394	$193	$7,527	$961.95	$539.236

Table 9: Descriptive Statistics for TL Tender Rate

We previously stated the findings for Spot and Spot_Dist during the down TL market in 2008-9. Using the same mathematical procedures we used to find the minimum cost for a load, we can calculate that the minimum cost for a spot load is $44.68 ($9.68 + 250*$0.14). For a minimum cost load ($607.31) this would mean that the cost would escalate to $652.00 and the spot surcharge would be 7% of that cost. In an industry as massive as the for-hire TL market, a 7% surcharge should not be ignored.

The finding of this model that spot rates are higher than contract rates adds more support to the assertion that aggregation can save companies money. The savings would come be avoiding the need to turn to the spot market because there are no contract rates established for a particular lane. Routing guides that only have point-to-point lanes are rarely collectively exhaustive and expose shippers to cost of unplanned freight as Harding showed (2005). Using lanes that are bid to regions enables companies to make their routing guides collectively exhaustive and more likely to avoid spot bidding. Therefore, including aggregated lanes in a company's routing guide may not only save money by lowering the premium associated with low-volume lanes, but it may also save money by avoiding spot premiums.

Using the model to determine whether to aggregate low-volume point-to-point lanes suggests that the region size at origin or destination would be relatively low. However, the model can also be used to assess whether using region in bidding could reduce the number of spot bids. A shipper should have a sense of where it is likely to have at least some spot bids, although the nature of spot bids suggests that shippers are unable to predict all the lanes that will require spot shipments. In our data set for example, there was a 6% repetition of spot bids in 2008 and 2009. In the event that a shipper can predict the areas where it is likely to have spot bids, it can use the model to assess whether it should bid a lane out with a region origin or destination to preclude the need for spot bids. However, if 94% of the spot volume would be new to the company's shipping system, as indicated by our data, then shippers would need to create a backup matrix of contract lanes to prevent paying spot premiums.

In order to use our model to assess whether to bid a lane with a region origin or destination, companies can perform a similar calculation to the one used above. However, the shipper would need more information about the shipment, specifically, the approximate length of haul. For example, if a shipper anticipates that it will have shipments from somewhere in the Detroit area to the Gary, Indiana area, a distance of about 250 miles, the minimum for a long-haul shipment, but the shipper does not know precisely where in the Detroit area the shipment will originate, they can assess the relative costs using this calculation:

$$\begin{aligned} &\mathit{Fixed\ Cost} + \mathit{Variable\ Cost} * \mathit{Distance} \\ &= \mathit{Premium\ per\ Mile\ at\ Origin} \\ &* \mathit{Average\ Distance\ from\ the\ Centroid} => \end{aligned}$$

$$\$9.68 + \$0.14 * 250 = \$0.50 * \sqrt{(\frac{A}{8})} => \frac{\$44.68}{\$0.50} = \sqrt{\left(\frac{A}{8}\right)} => 89.36 = \sqrt{\left(\frac{A}{8}\right)} =$$

$$> 7{,}985 = \frac{A}{8} => A = 63{,}882\ \mathit{square\ miles}$$

For this distance, a region origin of 63,882 square miles, or about 250 by 250 miles, will be more cost effective than the premium paid for a spot load. That is, in this example, a region the same size as the line-haul will yield a better price for the shipper than bidding the shipment on the spot market. As the distance gets larger, the "region size" will also grow larger, although it would exceed the boundaries of feasibility for a region. This corroborates what the data itself shows: a shipper has to pay more for a spot shipment than a contracted shipment.

So, while a breakpoint can be determined for when it is cost-effective to aggregate to a region-to-point or point-to-region bidding strategy rather than bid point-to-point, it will almost always be more cost-effective to establish a contracted region-to-region lane than to go to the spot market. It is our recommendation that shippers bid out overlapping lanes: point-to-point lanes that cover forecastable lanes used by the shipper, and region-to-region lanes to create a collectively exhaustive routing guide for lanes that have a historic pattern of spot lanes, since our data shows that it is cheaper to ship on a contracted lane than a spot lane, even during a buyer's market.

There are potential downsides to bidding regions to hedge against using the spot market. Adding in these additional lanes will increase the time it takes to prepare, bid out, and manage the routing guide. With just under 1,000 3-digit zip codes in the United States, covering all potential lanes with region-to-region bids would result in an unwieldy routing guide. Shippers should be selective in determining where they are likely to have spot shipments.

Even when aggregation might be appropriate, shippers should take care with how they choose to present this information. A comparison of identical 5 digit zip code pairings that were bid as point-to-point, point-to-region, and region-to-point show that the average price for point-to-point was over $50 less than for point-to-region. For this particular customer, Customer 7, there was a price break for lanes bid from region-to-point. According to our contact at CHR the region bid represented a specific set of origins within the region. Based on the information from him, we deduce that this type of region is similar to a bid of "City" meaning a specific origin or destination. That is, the carrier knows that while the origin is not a specific address, it in fact represents a single point or in the case of Customer 7's regions, a group of specific points, known both to the shipper and carrier. The shipper got a price break by aggregating. This seems to be sole scenario in the dataset where aggregation brought a lower price. Simply bidding out an origin or destination as a region to capture more loads on the broadly-defined lane does not itself result in cost-savings. The carrier does not have visibility on the points of origin or destination that will fall within the region origin or destination. Having this point information is what adds value for the carrier. This is most likely supportive of Mulqueen's (2006) thesis showing that carriers give discounts for loads terminating in a known origin point. Our example suggests that the more specific the information a carrier has about potential lanes, the better they are able to use that information within their network, driving down prices. Table 10 and Figure 13 give an example of one 5-digit zip pairing for Customer 7 that had contracted rates of four different types. In this case the point-to-point rate was roughly equal to one region-to-region lane, 3-digit zip to 3-digit zip but the contract rate for both of these was greater than the region-to-point lane. This point supports the finding of Hubbard (2001) that less than ideal information affects a carrier's bid for a lane.

43019 to 37743	City to City	State to City	3Zip to 3Zip	State to State	Total Shipments
Avg Cost Per Load	$ 632	$ 578	$ 616	$ 750	$ 621
Avg Cost Per Mile	$ 1.32	$ 1.21	$ 1.29	$ 1.57	$ 1.30
Total Loads	38	7	51	1	97
Length of Haul	478 Miles				

Table 10: Example of Lane with Various Bid Types

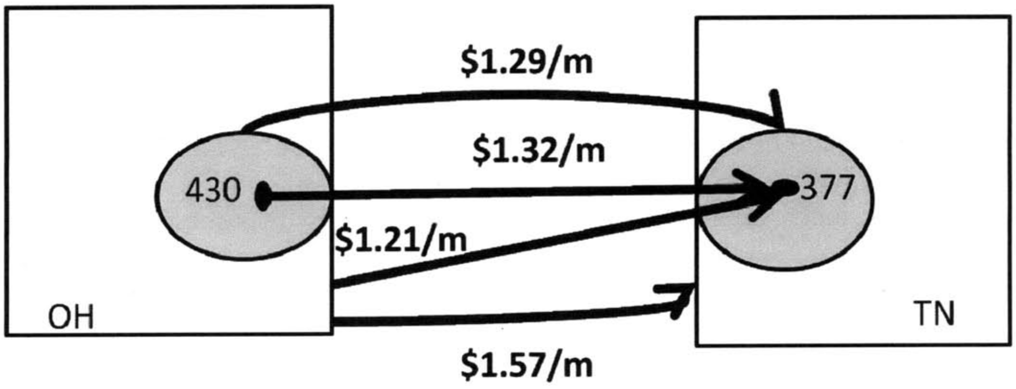


Figure 13: Example of Lane with Various Bid Types

6.3 Application of the model

The model is both simple and comprehensive enough that a broad range of shippers and carriers can use it to make better informed decisions about their aggregation strategies. While the model was developed using data from a transportation management system, the breadth and depth of the data set makes it applicable to users with smaller scope to their shipping networks. Shippers can use the model to perform a more critical analysis of the lanes that they bid out to carriers. They can assess potential cost savings of aggregation in their own networks quickly and easily. The more transportation costs impact a shipper's bottom line, the greater the potential benefit from cost savings from aggregation. Carriers can add value by being able to do a timely cost-benefit analysis of lanes bid out to them to assess whether a shipper's suggestion of aggregation will add value to their network. They can also add value to the shipper by offering aggregation of lanes in a shipper's routing guide if it would result in cost savings.

6.4 Further Research

Our thesis has also uncovered interesting information regarding spot market bids. Since our data spans a time period during which trucking companies did not have great bargaining power due to lack of business, further research should be conducted on spot bids for multiple time periods. Research focused on whether companies pay a premium for spot bids would be helpful to understand the cost of not having contract rates. If spot bids are consistently higher than contract rates then companies could use information in this thesis to ensure they have a collectively exhaustive routing guide. The competitiveness of spot bid relative to the length of a contract should also be investigated. If longer contracts or shorter contracts produce better results relative to the spot market, companies may also be better equipped to decide when to bid new contract out.

Since our data focused on aggregation's effect on contract rates we did not examine the truckload industry's brokerage effect on line-haul rates. While most of the spot bids in our data come from CH Robinson's brokerage business, we have not delved into all of the ramifications of getting a spot quote. Large variation may exist between differing truckload brokerage businesses that make further research important before making conclusions on spot quotes impact on truckload line-haul rates.

Reference List

Bertismas, D. and Freund, R. (2004) Data, Models, and Decisions: The Fundamentals of Management Science. Belmont, MA: Dynamic Ideas.

Bichler, M. and Steinberg, R. (2007) Introduction to the Special Issue on E-Auctions for Procurement Operations. *Production and Operations Management* v16 n4 (July-August): 401-403.

Caldwell, E., Fisher, B. (2008). The impact of lead time on truckload transportation rates. Masters of Engineering in Logistics Thesis, School of Engineering, MIT.

Caplice, C. (1996). An optimization based bidding process: A new framework for shipper-carrier relationships. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Caplice, C. (2006). Electronic markets for truckload transportation. Unpublished paper, Massachusetts Institute of Technology.

Caplice, C. and Y. Sheffi. (2003) Optimization-Based Procurement for Transportation Services. *The Journal of the Business Logistics*. v24 n2: 109-128.

Caplice, C. and Y. Sheffi. (2006) Combinatorial auctions for truckload transportation. In P. Cramton, Y. Shoham and R. Steinberg (Eds.), *Combinatorial auctions* (Chapter 21). Cambridge and London: MIT Press.

Chang, T. (2009) Decision support for truckload carriers in one-shot combinatorial auctions. *Transportation Research*. Part B n43: 522–541.

Chen, X., & Massachusetts Institute of Technology. Engineering Syst. (2003). Transportation service procurement using combinatorial. Masters of Engineering in Logistics Thesis, School of Engineering, MIT.

Chen, Y., S. Seshadri, E. Zemel. (2008) Sourcing Through Auctions and Audits. *Production And Operations Management* v17 n2, (March–April) 121–138.

Gujo, Oleg. (2008) Multi-attribute Inter-enterprise Exchange of Logistics Services. 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services.

Gujo, O., Schwind, M, Vykoukal, J. (2007). The Design of Incentives in a Combinatorial Exchange for Intra-Enterprise Logistics Services. In The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services.

Guo, Y, Lim, A, Rodrigues, B, & Zhu, Y. (2006). Carrier assignment models in transportation procurement. *The Journal of the Operational Research Society*, v57 n12: 1472-1481.

Harding, M. (2005). Can Shippers and Carriers Benefit from More Robust Transportation Planning Methodologies. Masters of Engineering in Logistics Thesis, School of Engineering, MIT.

Hubbard, Thomas N. (2001) Contractual Form and Market Thickness in Trucking. *The RAND Journal of Economics*, v32 n2 (Summer): 369-386.

IBISWorld. Ibisworld industry report: Long-distance freight trucking in the U.S. Technical report, January 2010.

Morgan Stanley Research North America (2008). Freight Transportation - Morgan Stanley Propriety Truckload Freight Index. April 4, 2008.

Sheffi, Y. (2004) Combinatorial Auctions in the Procurement of Transportation Services. *Interfaces* v34 n4 (July–August): 245–252.

Shmueli, G., Nitin, R., & Bruce, P. (2007). *Data Mining for Business Intelligence.* Hoboken: John Wiley & Sons, Inc.

Standard and Poor's Corporation. (2010) *Standard and Poor's Industry Surveys, Transportation: Commercial.* (7 January).

Zhelev, G. (2004). Flexibility in Transportation Procurement: A Real Options Approach. Masters of Engineering in Logistics Thesis, School of Engineering, MIT.

Appendix

A.1: Detail for Run 7 1

ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	131635132174.7	101.0	1303318140.3	49211.0	.0
	Residual	13220171247.1	499170.0	26484.3		
	Total	144855303421.8	499271.0			

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	318.7	1.2		267.4	.0
	Miles	1.1	.0	.8	1475.8	.0
	Spot	30.2	1.8	.0	16.5	.0
	Spot_Dist	.1	.0	.1	65.7	.0
	Inverse_Volume	26.5	2.2	.0	12.2	.0
	OriginEmptyMilesSquare	.9	.0	.0	36.9	.0
	DestEmptyMilesSquare	.5	.0	.0	39.1	.0
	O_AL	-81.2	1.4	.0	-56.7	.0
	O_AR	-95.8	11.4	.0	-8.4	.0
	O_AZ	-150.7	6.0	.0	-24.9	.0
	O_CA	-16.6	1.6	.0	-10.3	.0
	O_CO	-205.2	28.3	.0	-7.2	.0
	O_CT	-437.6	31.9	.0	-13.7	.0
	O_DC	-314.6	57.6	.0	-5.5	.0
	O_DE	-226.3	28.8	.0	-7.9	.0
	O_FL	-389.2	1.8	-.1	-220.7	.0
	O_GA	-108.0	1.4	.0	-78.7	.0
	O_IA	63.6	2.4	.0	27.0	.0
	O_ID	-716.6	94.0	.0	-7.6	.0
	O_IL	4.2	1.5	.0	2.8	.0
	O_IN	18.7	2.4	.0	7.6	.0
	O_KS	25.8	15.4	.0	1.7	.1
	O_KY	14.7	4.2	.0	3.5	.0
	O_LA	-203.6	3.2	.0	-63.1	.0
	O_MA	-520.9	1.8	-.1	-295.6	.0
	O_MD	-358.8	8.6	.0	-41.5	.0
	O_ME	-466.3	21.2	.0	-22.0	.0
	O_MI	-13.6	1.4	.0	-9.4	.0
	O_MN	181.8	1.9	.0	95.4	.0
	O_MO	-10.3	1.5	.0	-7.0	.0
	O_MS	-6.9	10.1	.0	-.7	.5
	O_MT	-449.6	115.2	.0	-3.9	.0
	O_NC	-47.6	1.0	.0	-47.4	.0
	O_ND	8.8	18.2	.0	.5	.6
	O_NE	67.0	19.7	.0	3.4	.0
	O_NH	-387.8	3.4	-.1	-113.0	.0
	O_NJ	-399.7	1.3	-.1	-297.5	.0
	O_NM	-213.8	162.8	.0	-1.3	.2
	O_NV	-107.0	2.1	.0	-51.8	.0
	O_NY	-349.1	3.3	.0	-105.8	.0
	O_OK	107.7	3.7	.0	29.2	.0
	O_OR	-300.7	2.9	.0	-103.8	.0

O_PA	-311.5	1.2	-.1	-270.6	.0
O_RI	-368.1	32.6	.0	-11.3	.0
O_SC	-27.9	2.7	.0	-10.5	.0
O_SD	148.9	18.6	.0	8.0	.0
O_TN	-93.4	1.3	.0	-70.9	.0
O_TX	-196.1	1.0	-.1	-201.2	.0
O_UT	-221.8	3.7	.0	-60.7	.0
O_VA	-273.5	3.8	.0	-72.7	.0
O_VT	-436.4	81.4	.0	-5.4	.0
O_WA	-108.8	3.5	.0	-31.3	.0
O_WI	59.7	1.4	.0	42.6	.0
O_WV	-130.5	12.2	.0	-10.7	.0
D_AL	-166.6	1.6	-.1	-105.5	.0
D_AR	-122.8	2.8	.0	-44.1	.0
D_AZ	63.7	2.0	.0	32.6	.0
D_CA	-246.3	1.6	-.1	-155.7	.0
D_CO	192.2	2.1	.0	92.4	.0
D_CT	340.0	3.7	.0	90.9	.0
D_DC	-400.6	115.1	.0	-3.5	.0
D_DE	194.7	10.6	.0	18.3	.0
D_FL	259.8	1.3	.1	204.0	.0
D_GA	-118.2	1.5	.0	-78.2	.0
D_IA	-239.4	1.6	-.1	-145.6	.0
D_ID	248.8	7.3	.0	34.3	.0
D_IL	-226.5	1.6	-.1	-145.9	.0
D_IN	-190.1	2.4	.0	-80.3	.0
D_KS	-177.7	2.5	.0	-71.1	.0
D_KY	-129.3	2.1	.0	-62.2	.0
D_LA	-72.5	1.9	.0	-37.8	.0
D_MA	421.7	2.1	.1	205.0	.0
D_MD	169.0	2.6	.0	65.6	.0
D_ME	510.7	2.7	.1	188.9	.0
D_MI	-98.1	2.0	.0	-48.0	.0
D_MN	-232.4	1.9	-.1	-120.2	.0
D_MO	-191.8	1.8	-.1	-106.8	.0
D_MS	-138.8	2.2	.0	-63.1	.0
D_MT	700.5	7.7	.0	90.7	.0
D_NC	-78.7	1.2	.0	-64.8	.0
D_ND	50.3	3.6	.0	13.8	.0
D_NE	-186.1	2.7	.0	-68.9	.0
D_NH	434.4	3.4	.1	128.0	.0
D_NJ	210.3	1.8	.1	118.8	.0
D_NM	223.3	3.4	.0	64.8	.0
D_NV	-98.7	2.6	.0	-38.3	.0
D_NY	269.1	1.7	.1	158.3	.0
D_OH	-105.3	1.3	.0	-79.9	.0

D_OK	-23.7	3.0	.0	-7.8	.0
D_OR	194.5	2.5	.0	77.3	.0
D_PA	104.8	1.3	.0	80.4	.0
D_RI	443.5	9.0	.0	49.3	.0
D_SC	-81.8	2.3	.0	-35.4	.0
D_SD	61.8	6.0	.0	10.4	.0
D_TN	-161.8	1.6	-.1	-101.0	.0
D_UT	251.5	2.5	.1	100.4	.0
D_VA	88.0	1.9	.0	46.7	.0
D_VT	498.6	12.0	.0	41.5	.0
D_WA	319.5	2.2	.1	142.7	.0
D_WI	-239.8	1.7	-.1	-144.3	.0
D_WV	1.5	4.1	.0	.4	.7
D_WY	494.6	4.5	.0	109.9	.0

A.2: Regression Runs Using Different Aggregation Methods

RUN Number	5	6	7	8	9	10	11	12	13	14
R-Squared	91%	91%	91%	91%	91%	91%	91%	91%	91%	91%
(Constant)	$ 323.91	$321.18	$322.31	$ 322.31	$322.40	$ 322.40	$ 56.42	$ 56.42	$ 319.63	$ 319.63
Miles	$ 1.14	$ 1.15	$ 1.14	$ 1.14	$ 1.14	$ 1.14	$ 1.15	$ 1.15	$ 1.15	$ 1.15
Spot	$ 13.71	$ 12.23	$ 9.68	$ 9.68	$ 9.40	$ 9.40	$ 3.99	$ 3.99	$ 3.25	$ 3.25
Spot_Dist	$ 0.14	$ 0.13	$ 0.14	$ 0.14	$ 0.14	$ 0.14	$ 0.13	$ 0.13	$ 0.13	$ 0.13
Inverse_Volume	$ 25.83	$ 42.24	$ 26.78	$ 26.78	$ 24.88	$ 24.88	$ 58.50	$ 58.50	$ 59.41	$ 59.41
O_Warehouse	$ 114.41									
O_3Zip	$ 40.38									
O_State	$ 75.71									
D_Warehouse	$ (106.46)									
D_5Zip	$ 68.20									
D_3Zip	$ 4.31									
D_State	$ 50.43									
Point_Region		$ 21.14								
Region_Point		$ (20.76)								
Region_Region		$ 69.80								
OriginEmptyMilesSquare			$ 0.91							
DestEmptyMilesSquare			$ 0.50							
OriginEmptyMilesCircle				$ 0.81						
DestEmptyMilesCircle				$ 0.44						
LaneEmptyMilesSquare					$ 0.60					
LaneEmptyMilesCircle						$ 0.54				
OrigLHRatioSquare							$ 0.17			
DestLHRatioSquare							$ 0.11			
OrigLHRatioCircle								$ 0.20		
DestLHRatioCircle								$ 0.13		
LaneLHRatioSquare									$ 0.06	
LaneLHRatioCircle										$ 0.07
Origin States Included	X	X	X	X	X	X	X	X	X	X
Destination States Included	X	X	X	X	X	X	X	X	X	X