GOING AWRY: UNDERSTANDING TRANSPORTATION BUDGET FAILURES

by

Venkateswara Rao Bandaru

Bachelor of Technology, Andhra University Visakhapatnam

and

Emilio Dolci

Bachelor of Management, University of Buenos Aires

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2020

Signature of Author: ____

Department of Supply Chain Management
May 15, 2020

Signature of Author: ____

Department of Supply Chain Management
May 15, 2020

Certified by: ____

Dr. Chris Caplice
Executive Director, Center for Transportation and Logistics
Capstone Advisor

Certified by: ____

Angela Acocella
Ph.D. Candidate, Center for Transportation and Logistics
Capstone Co-Advisor

Accepted by: ____

Dr. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

GOING AWRY: UNDERSTANDING TRANSPORTATION BUDGET FAILURES

by

Venkateswara Rao Bandaru

and

Emilio Dolci

Submitted to the Program in Supply Chain Management
on May 15, 2020, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

The US truckload transportation industry goes through phases of over and under supply of capacity, causing a dramatic impact on freight rates and transportation budgets. External factors like macroeconomic conditions, unexpected market forces, and changing regulatory policies tend to influence the velocity of these phases. We present an analysis of factors affecting the transportation budgets within the ambit of the transportation industry and shippers' procurement processes. The results of our research suggest that prevailing truckload market conditions impact shippers' transportation budget accuracy. The volume variation of a lane, and the origin or destination states, also have an impact, to a lesser degree. Higher awareness of the market conditions that influence transportation budget accuracy will allow shippers to be more effective in their planning processes.

Keywords: transportation, budget, procurement, freight rates, routing guide

Capstone Advisor: Dr. Chris Caplice
Title: Executive Director, MIT Center for Transportation and Logistics
Founder and Director, MIT Freight Lab

Capstone Co-Advisor: Ms. Angela Acocella
Title: Ph.D. Candidate, MIT Center for Transportation and Logistics

ACKNOWLEDGMENTS

We want to sincerely thank our advisor, Dr. Chris Caplice, and our co-advisor Angela Acocella, for their guidance, encouragement, support, and time given to us throughout the project.

We also want to thank our sponsor company C.H. Robinson and TMC, especially Steve Raetz, Andrew Welch, Mat Leo, Mark Meier, Fiorella Mete, and Ryan Gannon, for their valuable insights while analyzing the data and their expertise.

Next, we would like to thank Toby Gooly for her comprehensive review and patience throughout the writing process.

- *From both of us*

I would like to thank my wife, Vandana, and son Anirudh for their unconditional love and support. I am grateful for their patience and for allowing me to pursue my goals. I would also like to thank my parents for their encouragement.

My heartfelt thanks to my partner Emilio for his support throughout this project and a great sense of humor.

- *Venkat*

I would like to thank my family and specially my wife Malena for their encouragement and support; I could not have done it if it weren't for her. I am grateful for their love and the sacrifices they made to allow me to pursue this challenge.

My heartfelt thanks to my partner Venkat for his patience and hard work throughout this project.

- *Emilio*

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. Introduction

In 2018 the United States truckload market experienced some of its tightest conditions and highest rates in over a decade. These conditions had a significant impact on many shippers' overall financial performance. According to Truckstop.com data, spot market freight rates increased approximately 25 percent from the post-holiday seasonal trough in February to the summer peak, then fell approximately 20 percent through the end of the year. Economic activity in 2017 and 2018 led to excess demand for transportation as shippers rebuilt inventories in addition to adapting to growth in e-commerce and strong consumer demand. Tighter government regulations and rising driver wages have also contributed to higher freight rates (Bureau of Transportation Statistics, 2018). These market conditions and regulatory changes resulted in unbalanced capacity (supply) and demand (requested loads) and caused shippers across all industries to exceed their 2018 logistics budgets. The impact was so significant that many large shippers reported that freight had a material adverse effect on earnings (CSCMP, 2019).

The three objectives of the project are to find ways to estimate and generalize the frequency, impact, and nature of the transportation budget failures; to develop a probabilistic model to estimate the reliability of the budgets; and to identify key relevant managerial implications that could help increase the reliability of such budgets.

1.1. Overview of the trucking industry

According to American Trucking Associations (ATA, 2019), United States Business Logistics Costs (USBLC) rose 11.4 percent in 2008 over the previous year and reached 8.0% of the nation's GDP – a jump of 50 percentage points. The industry's trucks moved 11.49 billion tons of freight, 71.4% of the nation's freight tonnage. Trucking's revenues accounted for 80.3% of the nation's freight bill, jumping to nearly $800 billion in 2018, up from $700 billion in 2017. There were 7.8 million people employed in trucking-related jobs in 2018, an increase of 100,000 from the previous year (Bureau of Transportation Statistics, 2018). However, as discussed in Caplice and Sheffi (2003), economies of scale do not typically apply in the truckload (TL) industry, i.e., allocating more volume to a specific carrier does not always result in lower prices. The TL carrier's cost structure is more sensitive to economies of scope, where the cost of serving a lane depends on having an acceptable follow-on load. There is a high degree of uncertainty involved with securing a follow-on load, and carriers user pricing as a hedge against the uncertainty.

1.2. Market Structure

Despite its size, year-over-year growth, and the critical role it plays in the nation's economy, the trucking industry is extremely fragmented. The industry comprises hundreds of thousands of participants making buy-sell decisions based on a variety of market conditions, operating strategies, financial constraints, and other exogenous factors like economic cycles, regulatory changes, and weather events.

In the trucking industry, the demand is generated by hundreds of thousands of companies making decisions to move their freight between locations to serve their respective markets and customers. The supply side of the market is similarly broad and fragmented. According to the ATA, there were over 580,000 for-hire common carriers registered with the Federal Motor Carrier Safety Administration in 2015 and another 747,781 private carriers. Over 97 percent of those carriers reported operating fewer than 20 trucks, with almost 91 percent operating six or fewer. And while there are several large national carriers with tractor fleets in the 10,000–20,000 range, the top 15 carriers account for less than 12 percent of total for-hire market revenue. Therefore, this is a supply base that, in many ways, is dominated by the long tail of hundreds of thousands of small and medium-sized trucking companies and individual owner-operators.

Given the exceedingly high number of participants (both shippers and carriers) and low entry and exit barriers for carriers, no single participant or group of participants is large enough to dictate the market pricing. The trucking industry is also highly susceptible to economic activity, geopolitical conditions, demand shifts caused by the seasonality of goods, and significant weather events. Therefore, the stability of market pricing is very temporary and is mostly driven by the balance of supply versus demand and in a state of constant flux, as shown in Figure 1.

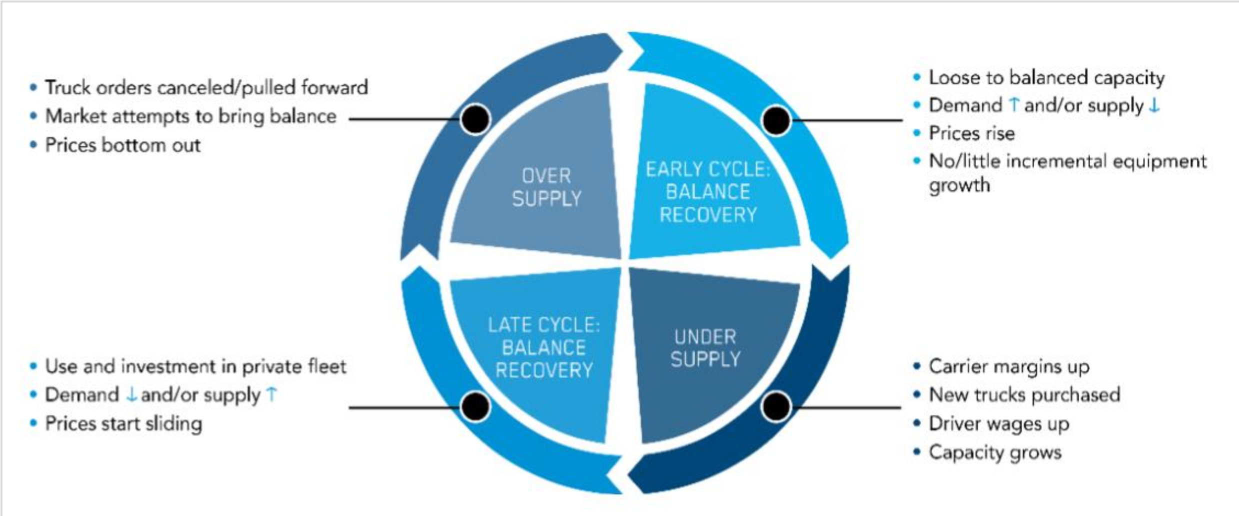

Figure 1: Truckload (TL) Industry Market Capacity Cycle (Source: C.H. Robinson, 2019)

As shown in the figure, the truckload (TL) market undergoes periods of oversupply and undersupply. During the oversupply phase, the market is flooded with more capacity, and competition among carriers drives prices down. This phase is also referred to as a 'Soft Market.' Similarly, during the undersupply phase, demand exceeds the available capacity. During this phase, carriers demand higher prices and reap higher margins. This phase is referred to as a 'Tight Market.' Between the oversupply and undersupply phases, there is a period of recovery where supply and demand tend to move towards equilibrium.

1.3. Other Factors Affecting Market Pricing

In addition to market cycles, factors like the seasonality of goods, demand shifts caused by significant weather events, and annual procurement cycles also play an essential role in influencing market pricing. Seasonal items experience a surge in demand over a short time in response to a planned production or demand window. Examples include harvest seasons in different regions, where speed to market to maximize shelf life causes inflationary rate conditions in outbound regions. Similarly, catastrophic weather events create shifts in both demand and supply in a relatively short period. These kinds of events not only generate a surge in demand but also cause disruptions to outbound freight, which results in procuring unplanned transportation capacity over irregular lanes. The broader network disruption acts as an additional inflationary force, thereby amplifying the market impact of weather events. In addition, distance accounts for 70-80% of the variability in transportation rates paid by shippers and rest are determined by factors such as regional sensitivity, dwell times, and freight imbalances (Caplice Class notes, 2012).

While market cycles, seasonality, weather events, and other factors cause industry wide price volatility, many organizations, especially those with high shipment volumes, employ an annual procurement exercise as a strategic lever to reduce their exposure to price volatility and supply risk. As part of the yearly procurement event, organizations go out to the market to set fixed contract rates for the next fiscal period. Different organizations adopt different procurement strategies and seek to set or reset contract rates for the lanes in their supply chain network to cover expected shipment volumes between nodes. The usual outcome of this annual procurement event is a routing guide, which helps organizations to establish and maintain primary and backup carrier relationships for lanes (Caplice 2009). Similarly, an unexpected surge in the shipper's freight volume may push the prices up in some market conditions. This unexpected surge in freight is also referred to as 'unplanned freight'. In summary, unplanned freight can happen in two scenarios., one, where unexpected surge volume on lanes that are part of bidding and included in the budget, and second, unexpected volume on lanes that are not part of bidding event and not included in the budget.

However, a unique aspect of the transportation contracts is that they are non-binding in terms of volume guarantees for both the carrier and the shipper. Primary carriers are expected to accept all loads tendered, but they can reject some loads without a direct penalty. Likewise, shippers are not penalized if all the volume promised in the annual bid does not materialize. If the primary carrier rejects the load, shippers will continue to move through the routing guide and tender the loads to backup carriers until an alternate carrier accepts the load. Tender rejections can happen for many reasons, including misalignment between shippers' and carriers' networks, market conditions, and the higher price differential between the contract rate and spot rates, among other reasons.

1.4. Impact on Transportation Budget

As discussed above, the trucking industry is highly susceptible to market cycles, trade policies, economic conditions, and seasonal and regional demand shifts, which, in turn, impact the transportation budget. For example, 2018 was a challenging year for shippers as capacity tightened, and rates rose. This volatility in pricing caused shippers across virtually all industries to exceed their 2018 logistics budget and had a significant negative material impact on earnings. Other factors such as quality of routing guide, primary carrier acceptance rate, the volume of unplanned freight, and expected volume vs. realized freight volume on lane could influence an organization's logistics budget.

1.5. Project Objective

The objective of the project is twofold. First, we aim to find ways to estimate and generalize the frequency, impact, and nature of unplanned freight and other factors on transportation budgets. Second, we develop a probabilistic model that will estimate the range of potential cost deviations from the planned budget. Unexpected freight costs arise due to the non-availability of primary carriers, higher volumes on lanes, the addition of new lanes, market cycles, and other macroeconomic factors.

The project sponsor is C.H. Robinson, a Fortune 500 provider of third-party logistics and multimodal transportation services. TMC, a division of C.H. Robinson, primarily operates on behalf of various shippers by offering a combination of global transportation management services (TMS) software, logistics process management, and consulting services.

In this project, we use transactional data from TMC to identify the factors affecting the transportation budget. We identify the frequency of these factors and their relative significance and impact on the transportation budget. Further, we analyze the freight procurement outcome data from TMC to understand how companies use procurement events to set up routing guides and to establish expected freight volumes, carrier relationships, and expected costs. We also utilize transactional data from TMC to explore the discrepancies between expected and actual freight volumes on lanes, acceptance rates by primary carriers, and the actual cost of transportation (line haul charges excluding fuel and other surcharges) to evaluate the real spend. We develop a statistical regression model to estimate the range of future expected costs (for the following year) by identifying patterns between estimated budget and actual spending.

The remainder of this report is organized as follows: Chapter 2 presents a literature overview of the problem; Chapter 3 discusses the data preparation, data characterization, and methodology used; Chapter 4 describes the regression models and their results; and Chapter 5 discusses managerial insights and recommendations for future research.

2. Literature Review

In general, transportation budget overrun can be attributed to routing guide failures and unplanned freight. We conducted a literature review of the existing body of work on both of these aspects. Our literature review is comprised of three main strands of literature:

- Broad truckload industry
- Routing guide failures and unplanned freight impact on transportation costs
- Methods to minimize transportation costs as a means of making more reliable budgets, irrespective of market cycles.

Truckload Industry

Council of Supply Chain Management Professionals (CSCMP) State of Logistics Report (2019) positions the 2019 market cycle in a period of contraction, coming from a very strong 2018 in terms of tight capacity and higher than usual rates. Many macroeconomic parameters, as well as industry-specific issues such as increased regulations, aging of the workforce, and an e-commerce surge, seem to indicate the beginning of a new upward cycle in the near future where budget reliability will be put to the test.

Caplice (1996) proposed a framework for the shipper-carrier relationship using optimization-based bidding and freight cost minimization strategies. Combinatorial auctions allow a carrier to submit a package bid for a set of lanes so that the carrier can take advantage of economies of scope. When a carrier can secure loads on multiple lanes, that ensures a continuous move and its overall cost per load decreases. The carriers pass on the benefit to the shippers in the form of a discount on the line haul rates (Caplice, 2009). Allowing carriers to capture economies of scope should lead to lower costs.

Routing Guide Failures

Bleggi and Zhou (2016) offered a strategic perspective on how to use carriers' attributes to establish a routing guide to improve the tender acceptance rate. They found that shippers that use leading carriers of the region or more focused carriers as their primary carriers tend to have a higher acceptance rate and better overall performance. In addition, according to research conducted by Yoo Joon Kim (2013), increasing weekly volume volatility is correlated with a higher tender rejection rate. The author found that tender rejections are not correlated with geographical patterns or the length of the haul. He also found that the differential between the primary rate and market rate did not explain tender rejections. However, a higher backup rate differential is associated with a higher rejection rate. These studies are important for us to understand the effect of strategies employed by shippers to identify and assign primary carriers. Understanding the routing guide failure patterns and their respective impact on transportation costs is vital to building a budget recommendation model.

Acocella, Caplice, Sheffi(2020) developed a method to determine the breakpoints and changes in freight market conditions. The method combines econometric methods and primary carrier acceptance ratio (PAR) as the measure of the market to determine the breakpoints. Knowing the prevailing market condition is vital to understand the behaviors of both shippers and carriers and to understand the performance of the transportation budget.

Transportation Costs

Aemireddy and Yuan (2019) calculated the impact of various shipment attributes, such as tender lead times, distance, regional sensitivity, lane consistency, and volatility, on the costs and tender rejection rates based on three years of historical data from TMC. They found that, among other attributes, 82% of the explanatory contribution comes from origin markets, destination markets, and the distance between certain origins and destinations. The authors also provide models for estimating a line haul rate to be paid to a carrier for a shipment, the likelihood of tender acceptance by the primary carrier, and the probability of a shipment ending up in the spot market. The regression models in the research show a strong negative correlation between lead time and primary carrier acceptance rate. This study is significant for us to understand the correlations between various shipment attributes and cost per load (CPL); incorporating some of these correlations is key to designing our budget recommendation model.

Similarly, 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 TL 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 its price. These studies are relevant for us to understand how the relationship between shippers and carriers and economies of scope, and how they influence transportation budget planning.

Collins and Quinlan (2010) argue that if a shipper has low-volume lanes, it should aggregate a group of such lanes into a region (less specific than 5-digit ZIP code) to increase the total volume the carrier can bid for in the auction process. Aggregating lanes can lower costs for shippers by avoiding spot market premiums and higher rates associated with less consistent lanes. However, the cost savings potential will depend on the size of the region, lane volume, and empty miles at the origin and destination. This insight is significant for us to extend the budget recommendation model to lanes with low and inconsistent volumes.

In addition to the line haul rates that shippers pay to carriers, every shipper also has its fuel surcharge (FSC) program. A carrier is usually compensated for any fuel costs when the price of the fuel exceeds a predetermined base price. Abramson and Sawant (2012) found that carriers determine the discount they may be willing to give on the line haul rate (cost per load) based on the fuel surcharge they can get. The more generous a shipper's FSC is, the higher the discount they get on line haul rates. Similarly, Caldwell and Fisher (2008) examined the effect of lead time on cost per load (CPL). They used distance, origin, and destination states, corridor volume, carrier size, tendered day, pickup day, and lead time as input variables for the regression model on cost per load. They showed that lead time could make a substantial difference to CPL and that longer lead times could lead to lower costs. These findings are very relevant for our research to understand the impact of shipper-carrier relationships, lane volume consistency, and other operational aspects on overall transportation costs and budget planning.

Zhelev (2004) provides a model for estimating contract rates vs. average spot rates for his data sample. The author finds that, on average, spot rates are about 20% higher than contract rates but can also be as much as 50–100% higher in some cases. Zhelev finds that the underlying factor that drives savings in TL transportation is the distribution of loads throughout the year.

Since we are building a budget recommendation model, the understanding of patterns in primary carrier rejections and the trade-offs between contract and spot rates, and prevailing market conditions are very important to our analysis. The transportation industry will continue going through market cycles, and shippers will continue to face challenges in managing transportation costs. The current body of work provides various models that shippers can implement to optimize the costs and analyze the impact of many operative parameters in the unplanned freight. However, we could not find any specific work related to budget reliability as a tool to guide shippers for better planning processes. We use four-year data from C.H. Robinson to estimate and generalize the frequency, impact, and nature of unplanned freight and routing guide failures on transportation budgets. This capstone project develops a probabilistic model to estimate the range of potential cost deviations from the planned budget.

3. Data and Methodology

In this project, we examine four years of truckload shipment data from 2015 Q3 to 2019 Q2. The data is provided by TMC (a division of C.H. Robinson) and includes all anonymized transactional data of its customers. The data consists of load transactions (physical movement of freight), tenders (tenders offered to carriers to move freight), and cost quotes (TMC's version of a routing guide). We adopted the following methodology to analyze the data and to draw useful insights, as shown in Figure 2 below.

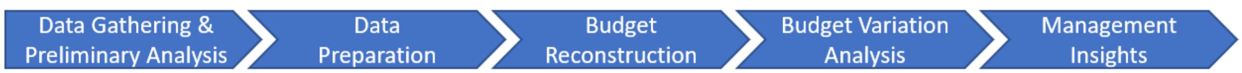

Figure 2: Data Analysis Methodology and Phases

In the 'Data Gathering & Preliminary Analysis' phase, we collect the data and conduct an exploratory analysis to understand the relationship between database entities, data dictionary, and scope of data. In the next phase, we identify data anomalies and exclude incorrect and invalid data. Data dictionaries of Loads data set and Tenders data set are added to Appendix A and Appendix B for reference. A list of business rules used for data cleanup is provided in Section 3.2. In the 'Budget Reconstruction' phase, we implement several business rules and methods to identify budget periods and recreate the shipper's transportation budget. In the 'Budget Variation Analysis' phase, we implement the regression models to test our hypotheses. And lastly, in the 'Management Insights' phase, we summarize the results, conclusions, managerial implications, and opportunities for future research.

3.1. Data Gathering & Preliminary Analysis

In this phase, we collect data from TMC and conduct an exploratory analysis to understand the data model, general characteristics, and key attributes. We also identify anomalies and establish general rules for data cleanup. The 'Loads' transaction data set has 1.3 million records; the 'Tenders' data set has 2.3 million records, and the 'Cost Quotes' data set has approximately 4 million records. A ‘Load’ data set record consists of load number (unique identifier), cost quote number (refers to a unique record in Routing Guide), an origin and destination addresses (city, state, and Zipcode), shipper company code, various timestamps of the transaction at origin and destination, and the charges (line haul, fuel, and accessorials) paid by the shipper, budget rate among other attributes. A ‘Tender’ data set record consists of cost quote number (refers to a unique record in Routing Guide), load number, tender method, carrier code (carrier identifier to which the tender is offered), sequence number (order in which the carrier is offered the tender), line haul, fuel and accessorial charges, shipper company code among other attributes. Similarly, ‘Cost Quote’ data set record consists of cost quote number (unique identifier), lane information (origin, destination), effective start and end dates of the quote, budget rate, primary carrier, and secondary carrier, among other attributes.

Using Tableau, we generate plots to observe the distribution of load transactions by the shipper, distribution of transactions across periods, number of distinct lanes (origin-destination), and line haul rate changes over a period of time. Figure 3 shows the volume of Load transactions by the shipper with at least 10k records. This process of data characterization helps us to understand the distribution of data better, correlations between various attributes, and to develop hypotheses.

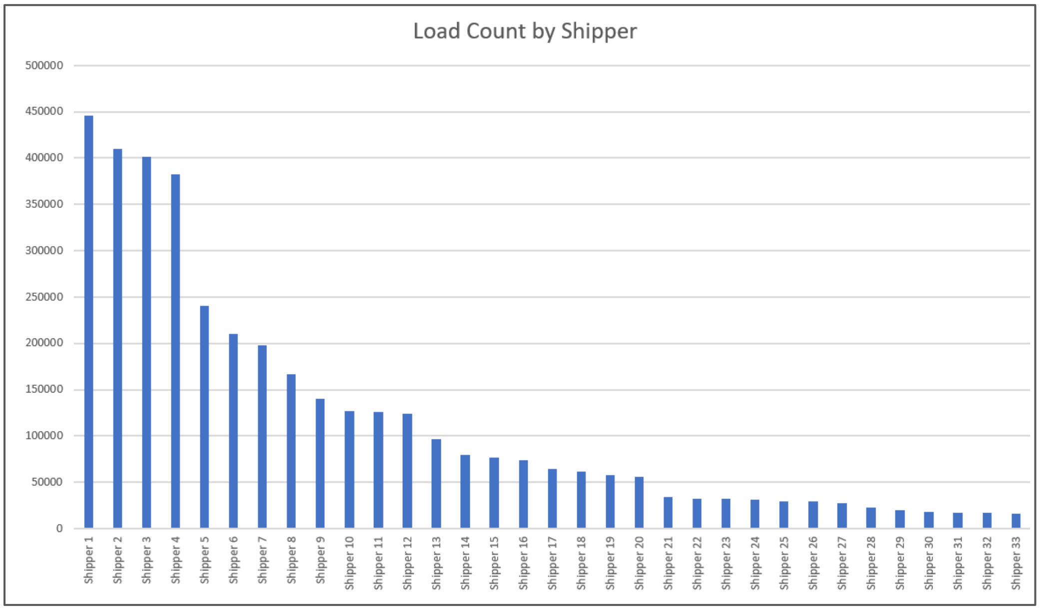

Figure 3: Load count by shipper with at least 10k load transactions

3.2. Data Preparation

In this phase, we mainly focus on preparing the data for analysis. Based on conversations with our sponsor company, we define business rules and clean up the data. We execute the following steps to clean up the data:

Data Cleaning Steps:

1. Select only relevant columns from Loads and Tenders data set to reduce the size.
2. Filter out records missing data in key columns like Load Number, Line Haul, Origin, Destinations, Carrier Name, Activity Date
3. Filter out records with negative Line Haul data.

4. Filter out Loads transactions with less than 250 miles distance.
5. Filter out Loads transactions where fuel charges are included in Line Haul.
6. Identify accepted tenders in the Tenders data set and filter out other records.
7. Join Loads data set and Tenders data set (accepted tenders) using the Load Number column. We need the effective start date of the load pricing from the Tenders data set.
8. Round the values of Line Haul and Lane Budget to the nearest dollar.
9. Create 'Origin City-Destination City' and Origin Zip-Destination Zip columns by concatenating Origin and Destination cities and ZIP codes.
10. Create Activity Week, Activity Month, Activity Year columns based on Activity Date column to classify transactions into yearly and weekly buckets.
11. Create a new column to store Lane Budget information for each transaction.
12. Create a new column to classify load transactions as 'Planned' or Unplanned. This classification process is explained in section 4.4.1.
13. Create a new column to classify load transactions as ‘On Budget,’ ‘OverBudget,’ or ‘UnderBudget.’ This classification process is explained in Section 4.4.2.
14. Create a 'Rate Variance' column to store the difference between Line Haul and Budget rates.

In addition to the above rules, we use other factors such as transaction volume, how long the shipper is using TMS application, the number of unique lanes (Origin City-Destination City/ Origin ZIP code –Destination ZIP code combinations), industry vertical, in choosing shippers for analysis. To avoid skewness of data, we consider only the top lanes, which cover up to 80% of total load volume and exclude the lanes with less volume. We do this because shippers tend to focus efforts on these lanes because it represents the majority of their business, and we expect most budget deviations to come from high volume lanes, as these have more likelihood of having a higher impact on the budget.

3.3. Budget Reconstruction

In this phase, we set the stage for building the regression models to test our hypotheses and determine the key factors affecting the shipper's transportation budget. To understand the variance in the transportation budget (planned vs. actual spend) for a given period, we need to establish the 'planned budget' of each lane for the period. There are two key steps involved in establishing the 'planned budget.' They are 1. Identification of the shipper's transportation main bid event and subsequent budget period; and 2. Determination of transportation budget rates for each lane that the shipper operates.

Transportation Budget Period

For most shippers, transportation budget planning and procurement processes follow a different cadence compared to traditional budgetary planning. The frequency of transportation procurement events is generally driven by shippers' business practices, changes in shippers' operating conditions, and transportation industry market conditions. Accordingly, the beginning of the period and the length of the period varies significantly.

Shippers conduct transportation procurement events periodically and invite bids from carriers for each lane they operate. The usual outcome of this annual procurement event is a routing guide, which helps shippers to establish and maintain primary and backup carrier relationships for lanes (Caplice 2009). The routing guide also contains the negotiated rate information and effective dates of the rates. In addition to period procurement events, sometimes shippers also conduct smaller procurement events (also knows as mini-bids) in response to changes in market conditions as well as changes in the shipper's business operations. TMC stores and maintains its customers' routing guide information in its TMS application.

The primary source for routing guide information for our project is the 'Cost Quotes' data set. This data set is supposed to have routing guide details for the customers of TMC (represented by the Shipper column in the data set). However, since TMC services are directed by each customer's individual operations, the 'Cost Quote' dataset must be supplemented with additional analysis to create consistent procurement events and lane budget identification across all customers.

To deal with high heterogeneity and missing data (in some cases), we adopt a simple statistical method to identify potential bidding events. The method involves classifying load transactions into weekly buckets and counting the number of primary carrier rate changes each week. We use the mean and standard deviation of primary carrier rate changes to identify weeks with the highest number of rate changes relative to neighboring weeks. Since procurement events usually run for several days, we consider a block of time (2–3 weeks) to identify potential procurement events. We mark these blocks of times as 'Bidding Events' and the beginning of 'Budget Periods.' We are able to identify the periodic bidding event pattern for many shippers. However, this approach has limitations since we rely on transactional data rather than on award data. As shown in Figure 4, in cases where a clear pattern emerged, we are able to identify the beginning and end of each budget period successfully. In other cases where there is no clear pattern, we considered 52 weeks (one full year) as the length of each budget period.

In addition, we conduct a survey with the account managers to confirm the procurement event dates we identify and made changes accordingly where such information is available. Figure 4 shows an example of procurement events for a single shipper. The data suggests that there is a procurement event in April/May every year, where we noticed a higher number of lanes are assigned new primary carrier rates.

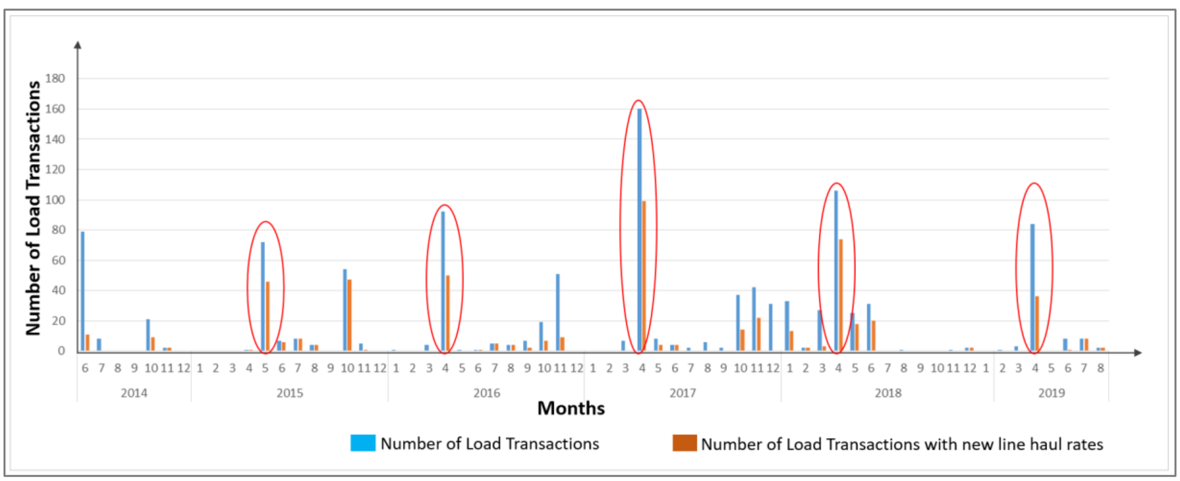

Figure 4: Procurement events as suggested by data for a shipper

Transportation Budget Rates

A 'Budget Rate' is the load price for a lane (origin-destination) that the shipper believes will be representative of the lane cost during the budget period. The 'Budget Rate' is usually estimated by the shipper based on the outcomes of transportation procurement events. For strategic reasons, sometimes shippers may choose to have more than one primary carrier with different rates, and the 'Budget Rate' is an average of those different rates. The level of specificity of origination and destination points of a lane can vary. Budget rates can vary considerably depending on the level of specificity. In the hierarchy of specificity, at the highest level is the actual addresses of both origin and destination, followed by 5 digit Zipcode – 5 digit Zipcode, City – City, 3 digit Zipcode – 3 digit Zip Code, and State - State. In our data set, we observe that many shippers use different levels of specificity for origin and destination. In our analysis, we create the origin-destination pairs using the hierarchy of specificity as presented in the routing guide without any modifications.

We also observe that in some instances, only a few lanes are part of all bidding events, while budget rates for the rest of the lanes remain active without any revision for a long time (as long as 36 months). These peculiarities pose additional challenges for determining how long the budget rates are valid and which lanes are part of the bidding process.

To mitigate these challenges, we use the 'Tenders' data set to derive budget rates from the transactions awarded to primary carriers at the beginning of each budget period. In other cases, we use the 'Loads' data set to derive budget rates for the lanes from the transactions within the first 30 days from the beginning of each budget period. However, these mitigations have limitations since we are entirely relying only on the transactional data and not on the Routing Guide data.

Also, we use the following business rules to treat the data uniformly across different periods and shippers:

a. There is no carry forward of the budget rate across budget periods. If a lane is missing from the bidding event of the following period, we do not carry forward the rate from the previous year.
b. If there is a clear pattern of bidding events emerged, then the budget rate for the period is valid for the entire budget period, irrespective of the length of the period. The length of the budget period varies from 5 months to longer than 24 months across shippers.
c. If a bidding period's end date is not well defined, then the budget rate is valid only up to 52 weeks (one full year).

In summary, due to the inconsistent and asymmetric nature of data across shippers, we are not able to directly use routing guides to identify budget rates. Alternatively, we used transactional data as an alternative source to derive bidding events based on what the data suggests. The next step in the process is to identify suitable shippers and lanes for analysis.

Shipper Selection for Budget Reconstruction

We analyze the combined data set (Loads + Tenders) and identify the shippers that have the availability of budget rates for a high percentage of lanes. Table 1 shows the top shippers that have budget rates (from Tenders data set) available. As shown, the percentage of availability of budget rates quickly drops. Shippers highlighted in color have budget information for at least 50% of transactions.

Table 1: Top 15 shippers with budget rate availability

Shipper	Count of Load Transactions	# of Load Transactions with Budget data	# of Load Transactions without Budget data	Percentage of Lane Budget Availability
Shipper 19	58083	44126	13957	75.97%
Shipper 7	197652	140335	57317	71.00%
Shipper 1	445737	314977	130760	70.66%
Shipper 6	210518	126060	84458	59.88%
Shipper 33	16441	8666	7775	52.71%
Shipper 4	382926	201197	181729	52.54%
Shipper 3	401304	206830	194474	51.54%
Shipper 8	166590	54996	111594	33.01%
Shipper 41	5999	1854	4145	30.91%
Shipper 28	22452	6211	16241	27.66%
Shipper 9	140422	34694	105728	24.71%
Shipper 15	76678	14347	62331	18.71%
Shipper 21	34007	4658	29349	13.70%
Shipper 27	27093	3510	23583	12.96%
Shipper 23	32048	3780	28268	11.79%

We identify a subset of shippers that follows a periodic procurement event process with an identifiable beginning and end of budget periods. However, each shipper follows a different cadence, and the duration of each period varies. Figure 5 shows a subset of shippers with budget periods plotted against the timeline (from 2015-2019).

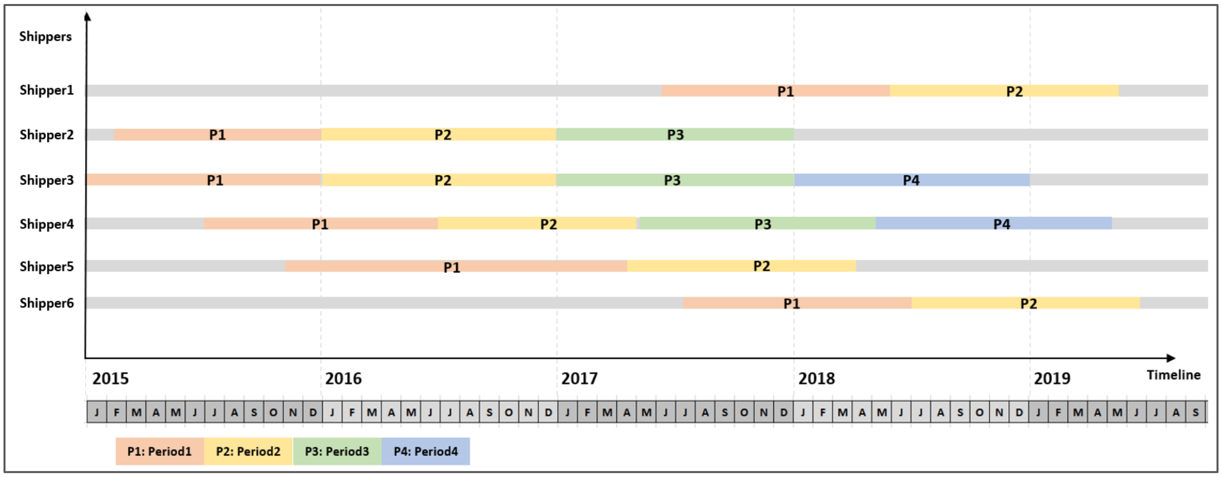

Figure 5: Budget periods of selected Shippers

In summary, we are able to find a few suitable shippers by applying business rules established to identify budget periods. The next step is to identify suitable lanes within each of these shippers.

Lane Selection for Budget Reconstruction

We observe that for the majority of the shippers, the top 80% of the load volume belongs to less than 20% of lanes. The remaining 20% of load volume is distributed across a very high number of lanes with very low volumes. Figure 6 shows a sample of shippers and the distribution of distinct lanes and load volume.

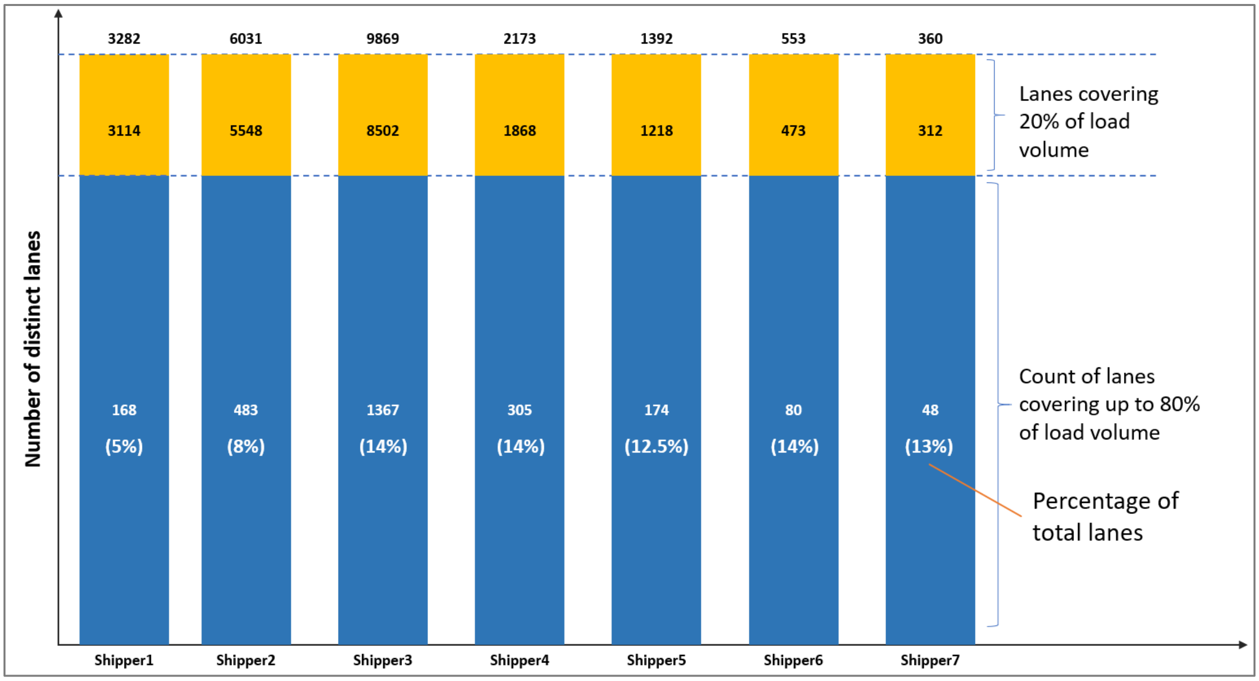

Figure 6: Number of distinct lanes covering up to 80% of load transactions vs. total lanes

The above figure shows the distribution of lanes and the volume of load transactions for a sample of shippers with known bid events. For example, shipper 3 has 9,869 distinct lanes, of which 1,367 lanes (14%) contributed 80% of the load transaction volume. We keep only the lanes contributing to the 80% of load transaction volume and exclude the remaining to maintain consistency in budget variance analysis. Also, the lanes represent most of the business that these shippers typically focus their attention on.

At the end of this step in our methodology, we extract a subset of data that has met all conditions of the selection criterion. Each of the seven shippers in the data set has well-defined bidding events and budget periods, availability of budget rates for a high percentage of lanes, and lanes with reasonable load volume across budget periods. The next step in our analysis is to build regression models to test our hypotheses and budget performance.

3.4. Budget Variance Analysis

In this step, we build regression models to identify the factors that affect the shipper's transportation budget and cause a variance in planned vs. actual spending. There are three steps involved in preparing the data before regression models can be implemented. First, we need to classify each budget period as 'Soft Market' or 'Tight Market' based on prevailing market conditions. Second, classify each lane as 'Planned' or 'Unplanned. Third, classify load transactions based on the variance between the planned budget rate and line haul (the actual price paid by the shipper).

Budget Period Classification Based on Market Conditions

As mentioned in section 1.2, the truckload industry market goes through periods of over and under supply. Acocella, Caplice, Sheffi(2020) identified the breakpoint of the last major transition of the market as 2017 Q2. Figure 7 shows the market transitions and the behavior of spot and contract rates. In our analysis, we used the breakpoints as a reference to classify shippers' budget periods. We classified the budget period based on when the main bid event that we assume is the same time when the budget is created. For example, as shown in Figure 5, the budget period P3 of Shipper 2 and Shipper 3 spanned across 2017 Q2, and we classified the market condition for these periods as 'Tight Market.' In these cases, even though the bidding events happened solidly in 'Soft Market,' the significant portion of execution happened after 2017 Q2, and therefore we classified the entire period as 'Tight Market.'

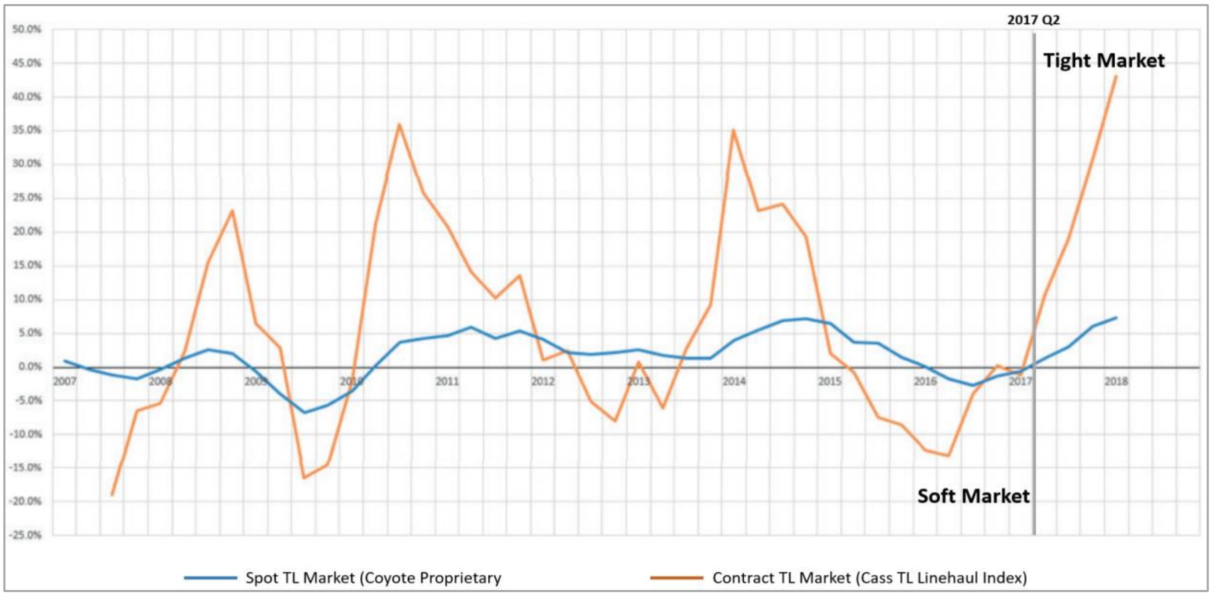

Figure 7: Spot vs. Contract TL rate behavior and market transitions (Source: Pickett C, 2018)

Lane Classification Based on Bidding event

A lane (Origin-Destination) is considered 'Planned' if the lane is part of a main procurement event, a primary carrier is assigned, and the rate is known. Similarly, a lane is considered 'Unplanned' if the lane (Origin-Destination) is not part of a main procurement event. In this case, the lane could be an existing one or a new addition to the shipper's network. Subjective factors like strategic importance, the complexity of a supply chain network, expected load volume, market conditions, and the relationship between shippers and carriers play an important role in deciding which lanes are included in the bidding event. In our data set, we observed that the percentage of lanes included in the bidding event varied across shippers. Table 2 describes the lane classification logic.

Table 2: Classification of Load lanes based on participation in the bidding event

#	Lane in the bidding event	Budget rate available	Classification1	Comments
1	Yes	Yes	Planned	
2	No	Yes	Unplanned	Budget rates are available but added to the data set in dates outside of the bidding event window, possibly via a mini-bid event
3	No	No	Unplanned	

Figure 8 shows an example of the load classification between Planned or Unplanned

Figure 9 shows an example of the load classification between Planned or Unplanned for one shipper. The budget performance degradation with time can be seen. As the time after the main bidding event progresses, more unplanned lanes are used. After each bidding event, the proportion of loads in planned lanes increases.

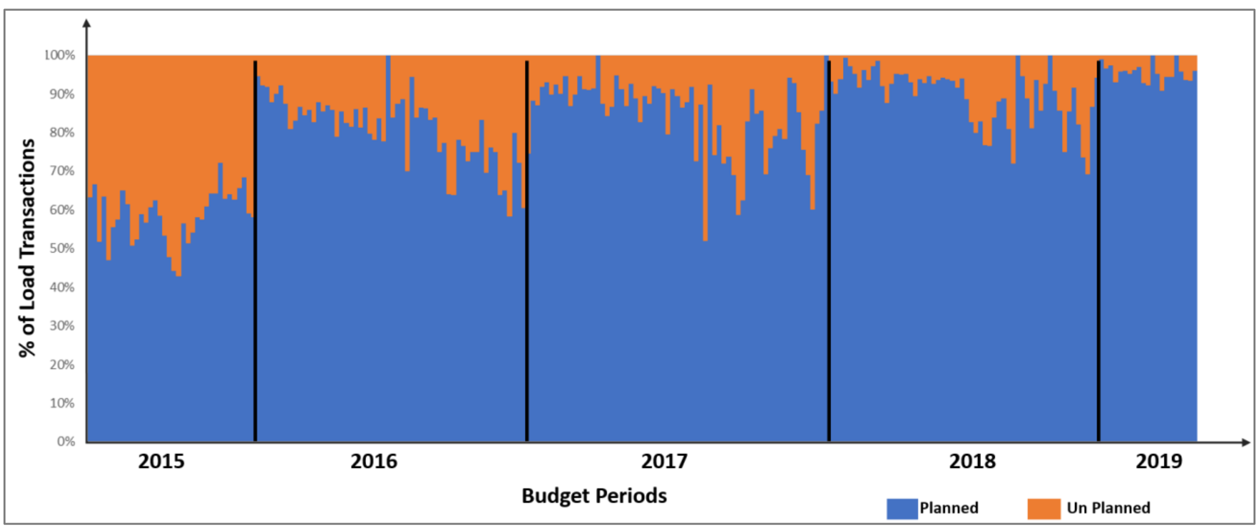

Figure 9: Planned vs. Unplanned weekly load % of a shipper
(black vertical lines indicate main bidding events)

Lane Classification Based on Spend Variance

The next step is to further classify each planned lane for each period based on how much shippers paid for each load transaction compare to the budget rate. The spend variance of load transactions is defined as the difference between 'Budget Rate' and 'Line Haul' (actual amount paid by the shipper excluding fuel, accessorials, and other surcharges). It is summarized at the lane and period level and stored in the 'Rate Variance' column in the data set. Figure 10 shows the distribution of the budget variation percentage for all lane-period combination of the entire dataset.

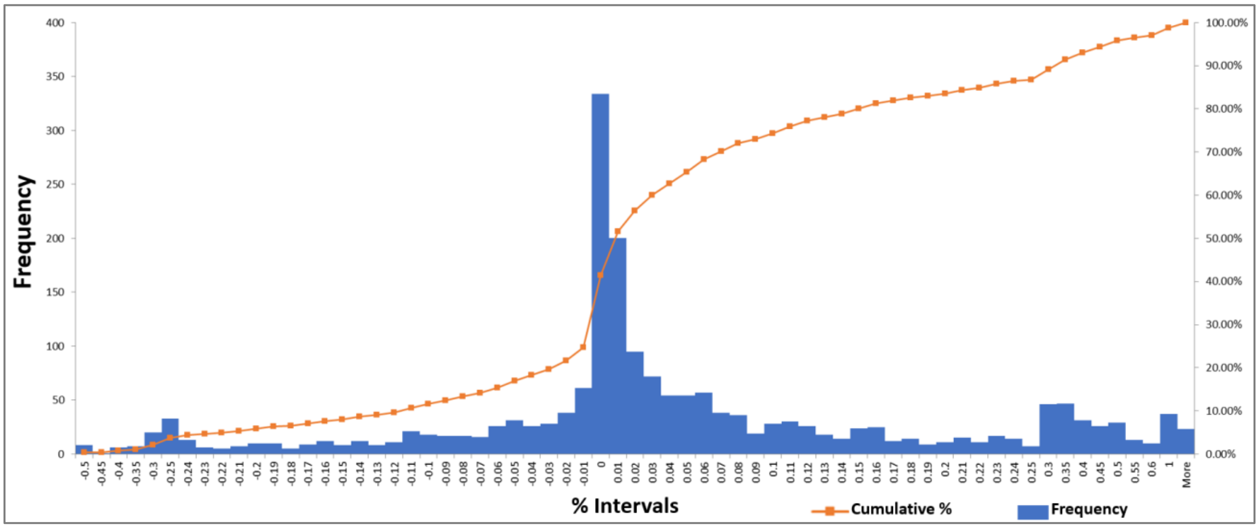

Figure 10: Budget variation distribution

Each lane-period is classified into two main categories:

- 'On or Under Budget': all lanes with negative rate variance (Under Budget) or zero (On Budget). We also include all lanes with a positive variance of up to +2% of the budget rate.

- ‘Over Budget’: all lanes with positive variances at least +2% higher than the budget rate.

This classification reflects the business approach to budget variation analysis. Shippers typically focus their attention on the lanes that show relevant budget overruns. For this analysis, we selected a 2 percent as a threshold level based on the sponsor company’s input. Table 3 describes the load classification logic.

Table 3: Classification of Load transactions based on Rate Variance

#	Lane Classification	Rate Variance	Load Classification
1	Planned	Line Haul – Budget Rate	‘On or Under Budget’: if Rate Variance < +2%) ‘Over Budget’: if Rate Variance > +2%
2	Unplanned	Line Haul	‘N/A’

As explained above, for the regression analysis, we combine the ‘Under Budget,’ and ‘On Budget’ plus 2 percent tolerance categories, but we used the full range classification for the preliminary analysis. Figure 11 shows an example of the load classification for a shipper between ‘On Budget’, and ‘Over Budget’. The budget performance degradation with time can also be seen in this case. As the time after the main bidding event progresses, a higher level of ‘Over Budget’ loads can be observed. After each bidding event, the proportion of ‘On Budget’ or ‘Under Budget’ loads increases.

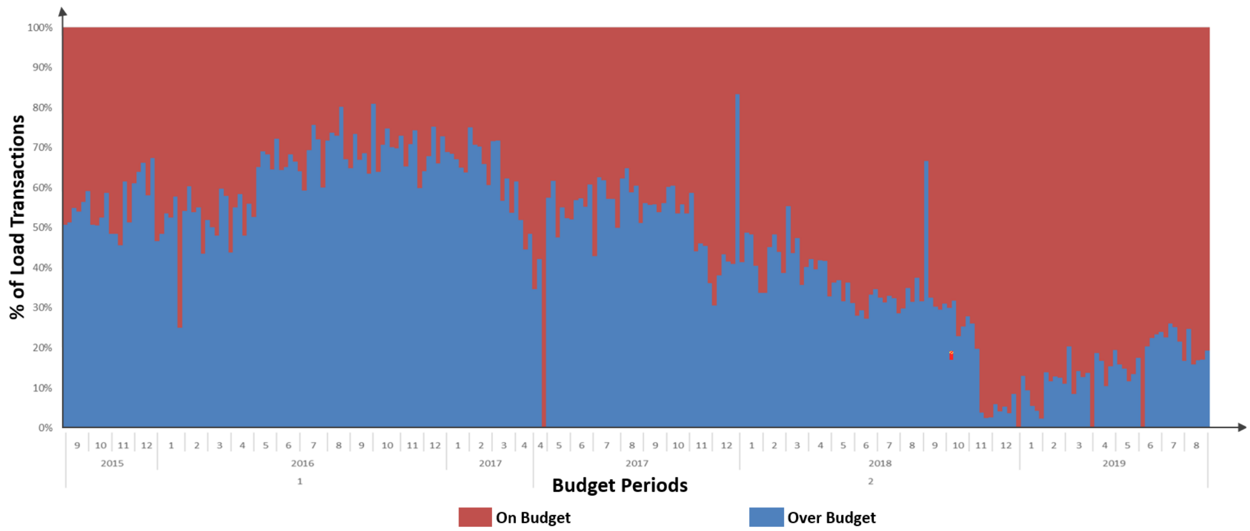

Figure 11: Budget performance weekly load distribution of a shipper

Regression Analysis

Regression analysis is a statistical technique for studying linear relationships. It begins by supposing a general form for the relationship, known as the regression model:

$$Y = \alpha + \beta 1X1 + ... + \beta kXk + \varepsilon$$

Y is the dependent variable, representing a quantity that varies from observation to observation (in our case, transaction to transaction), and is the primary focus of interest. X_1, ..., X_k are the explanatory variables (also known as ‘independent variables’), which also vary from one observation to the next, and are thought to be related to Y. Finally, ε is the residual term, which represents the composite effect of all other types of individual differences not explicitly identified in the model (Shmueli, Patel, Bruce, & Torgo, 2017). The objective of this project is to capture how different attributes of load transactions, Routing Guide, and exogenous factors like transportation industry market conditions impact shippers’ budgets; hence multiple regression models are built.

Ordinary Least Squares (OLS) multiple linear regression is used to model the relationship between the shipper's budget performance measured as percentage variation between the budget rate and the actual rate summarized at lane level (dependent variable) and various attributes (independent variables) of each lane and budget periods. The attributes included in the model are the lane's origin state, its destination state, budget period length, lane average monthly loads, lane coefficient of variation of monthly load volume, the distance between origin and destination, and the shipper's industry vertical.

The primary result of regression analysis is a set of estimates of the regression coefficients α, β1..., βk. These estimates are made by finding values for the coefficients that make the average residual 0, and the standard deviation of the residual term as small as possible. The result is summarized in the prediction equation: $Y_{pred} = a + b_1X_1 + ... + b_kX_k$ (Shmueli, Patel, Bruce, & Torgo, 2017).

Logistic regression is used to determine the probability of overall transportation cost staying within the budget given the transportation market conditions (Soft vs. Tight), the lane's origin state, its destination state, lane budget period length, lane average monthly loads, coefficient of variation of monthly load volume, the distance between origin and destination, and the shipper's industry vertical.

A detailed explanation of each regression model is provided in section 4 Results and Discussion.

Regression Parameters

In order to select the final parameters for the models, we regressed several lanes and shipper characteristics and tested our initial hypotheses. The final models are created using statistically significant parameters only. Table 4 shows all the preliminary characteristics of each lane or shipper and the preliminary hypothesis that we tested:

Table 4: Parameters for Budget Accuracy Regression Models

Parameter	Description	Hypothesis
Mileage	Distance between origin and destination points (min: 250 mi.)	Budget accuracy is negatively correlated with mileage
Bidding Events	Quantity of secondary bidding events (or "mini-bids") per lane	Budget accuracy is negatively correlated with bidding events. Re bidding is used to mitigate or corrects the impact of budget variations
Industry	Industry vertical of the shipper (for all the lanes of the shipper). The general category of the economy or the sector where the shipper operates, i.e., Automotive, Manufacturing, or Paper.	Budget accuracy is correlated with the shipper industry, i.e., certain industries may see more budget overruns than others
Coefficient of Variation	Level of the variability of the weekly volume (volume standard deviation/volume average) per each lane segmented by relevant intervals	Budget accuracy is negatively correlated with the coefficient of variation. Volume volatility increases the likelihood of budget overruns
Shipper	Shipper unique identification code	Budget accuracy is correlated with the shipper ID
Average monthly spend	Amount of dollars spent on average per month on each origin-destination lane	Budget accuracy is positively correlated with the average monthly spend
Average monthly volume	Number of loads shipped on average per month on each origin-destination lane	Budget accuracy is positively correlated with the average monthly volume
Average load price	Average price per load for each origin-destination lane	Budget accuracy is positively correlated with the average load price
Origin State	State where the shipment originated	Budget accuracy is correlated with some origin states
Destination State	State where the shipment is delivered	Budget accuracy is correlated with some destination states
Market Condition	State of the transportation market, defined as "soft" or "tight."	Budget accuracy is correlated with the market condition

For the coefficient of variation parameter (CV), we did some further analysis to identify relevant intervals to use in the regressions instead of using it as a continuous variable. We segmented the CV to capture the relevant intervals, as shown in Figure 12:

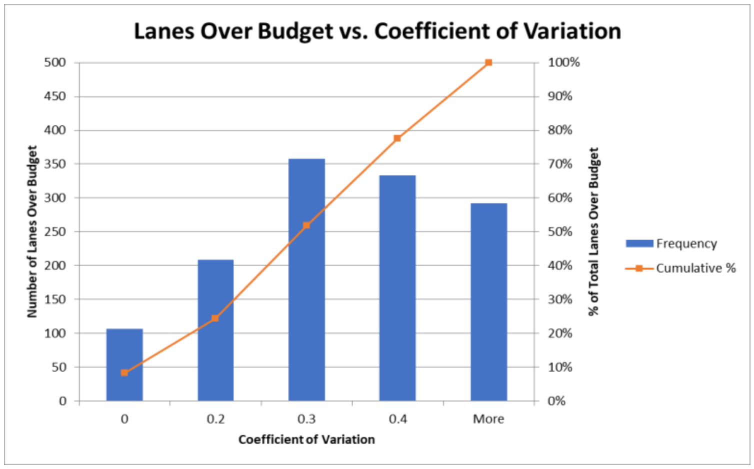

Figure 12: Budget performance weekly load distribution

4. Results and Discussion

In this section, we present the results from the quantitative analysis of the data. We use linear regression and logistic regression models to identify relationships and quantify the impact of different parameters on the probability of budget overruns when that budget is created.

We find statistically significant relations with five input variables, namely Market Condition, Coefficient of Variation, Origin State, Destination state, Quantity of Bidding Events during the budget period.

Linear Regression Model

The first model we use is linear regression, and the results and the impact of each parameter are shown in Table 5:

Table 5: Linear Regression Model Results **(Dependent Variable: Budget variation percentage (negative=savings, positive=budget overruns)**

Summary of Fit	
RSquare	0.0327
RSquare Adj	0.0278
Root Mean Square Error	1.0522
Mean of Response	0.0985
Observations (or Sum Wgts)	1986

Parameter Estimates		
Term	Estimate	Prob>|t|
Intercept	0.0152	0.5954
Market[Soft]	-0.0622	0.0219
CV Interval[>0.4]	0.0511	0.2385
CV Interval[0 to 0.2]	0.0251	0.5383
CV Interval[0.2 to 0.3]	-0.0561	0.1563
O-AZ	0.8774	<.0001
O-IA	0.5982	0.0002
O-NV	-0.1753	0.3478
D-CA	0.3448	0.0026
D-OR	-0.4142	0.1276
D-PA	0.2255	0.031

This model has a low R square; therefore, the level of explanation is low, but we use it to identify the origin and destination states with the highest impact. For the states with high statistical significance, we can use the sign of the estimate to identify the direction of the impact in terms of higher probabilities of overruns (positive signs) and savings versus the budget (negative signs). We include all origin and destination states in the model but excluded the states with very low statistical significance. Only the origin and destination states with significant p-values are included in the model.

Logistic Regression Model

For the second model, we use logistic regression, and the results are shown, and the impact of each parameter are shown in Table 6:

Table 6: Logistic Regression Model Results **(Dependent Variable: Lanes classified as On or Under Budget vs. Over Budget)**

Summary of Fit	
RSquare (U)	0.0609
AICc	2520.86
BIC	2548.8
Observations (or Sum Wgts)	1986

Parameter Estimates		
Term	Estimate	Prob>ChiSq
Intercept	0.7541	<.0001
Market[Soft]	0.6744	<.0001
CV Interval[0 to 0.2]	0.1993	0.017
CV Interval[0.2 to 0.3]	0.1719	0.0307
CV Interval[>0.4]	-0.2887	0.0007

This model also has a low R square but higher than the linear regression. We use this model to identify the probabilities of budget overrun.

To characterize the bidding event parameter impact, we used a separate model since the parameter is not usable as a predictor, and it is only known after the fact. In any case, the characterization of the secondary bid events could lead to other managerial implications and can be further analyzed in the future with additional data. For the third model, we also use logistic regression for the bidding event, and the results are shown, and the impact of the bidding event parameter are shown in Table 7:

Table 7: Logistic Regression Model Results – Quantity of Bidding Event (Dependent Variable: Lanes classified as On or Under Budget vs. Over Budget)

Summary of Fit	
RSquare (U)	0.0535
AICc	2534.75
BIC	2545.94
Observations (or Sum Wgts)	1986

Parameter Estimates		
Term	Estimate	Prob>ChiSq
Intercept	1.299	<.0001
Q_Bidding_Event	-0.564	<.0001

This model also has a low R square, but the statistical significance of the quantity of bids parameter is very high.

4.1. Market Condition

We used a logistic regression model to estimate the market condition impact. Market condition is the most significant indicator of the probability of a budget overrun. The probability of going overbudget is almost four times less likely in a soft market. Similarly, there is a 2.5 times probability of going over budget when the market transition happens from the soft market to a tight market. We used logistic nominal regression to estimate the probabilities of budget overrun based on the market conditions at the time the budget is issued. The results of the model are shown in Table 8:

Table 8: Probability of Budget Overrun vs. Market Condition

Market Condition	On or Under Budget	Over Budget
Soft Market	81%	19%
Tight Market	52%	48%

Even if the probability of budget overrun in tight market conditions is significantly higher than during soft market conditions, the combined effect with other parameters will increase the probability of overrun even further.

4.2. Coefficient of Variation

We use a logistic regression model to estimate the volume variability impact. High lane-volume variation is the second-highest indicator of potential budget overruns. High levels of lane volume variability increase the probabilities of budget overruns. The results of the model are shown in Figure 13, where we plot the probability of a budget overrun for the range of CV values for tight market budget periods (orange line) and soft market budget periods (blue line):

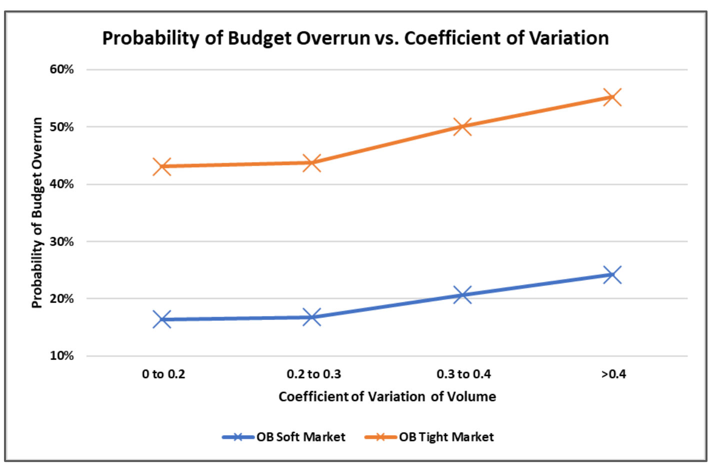

Figure 13: Probability of Budget Overrun vs. Coefficient of Variation

For example, a lane with a CV of .3 to .4 and a budget set during a tight market has about a 50 percent chance of going over budget. The probability of budget overrun developed during a tight market condition for a lane with high volatility (i.e., CV>0.4) is 55 percent, compared with 44 percent in a low-variability scenario (i.e., 0 <=CV<=0.2). Lanes with inconsistent volume will require a higher degree of attention during the budgeting process to reduce the possibilities of budget overrun.

4.3. Origin and Destination States

Some geographic areas have a material effect on budget performance. We used Ordinary Least Squares (multiple linear regression to identify the impact of specific origin/destination states on the budget overrun percentage. The analysis shows that some states have a significant impact on the percentage of variation of the budget overrun.

Even if some origin and destination states have a high degree of statistical significance, the low level of R-squared ($R2=0.0327$) of the model does not allow the extrapolation of consistent probabilities. We include all origin and destination states in the model but excluded the states with very low statistical significance. Only the origin and destination states with significant p-values are included in the model, as shown in Figure 14 (O-state indicates origins while D-state indicates destinations).

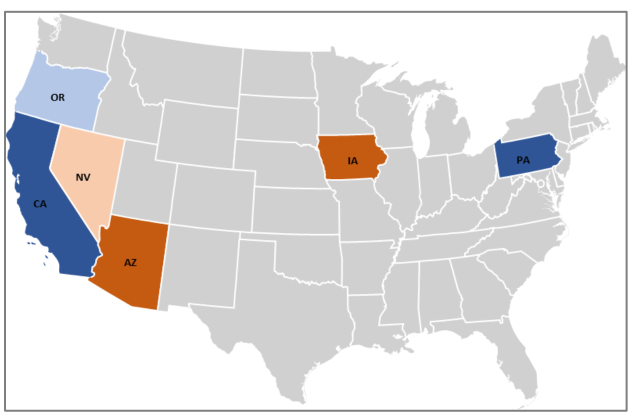

Figure 14: Destination states (in blue) and Origin states (in orange)

(Darker colors indicate states with higher statistical significance)

More data will be needed to estimate the probable impact of a budget overrun of a lane based on the origin or destination state.

4.4. Bidding Events

We used a logistic regression model to estimate the bidding events parameter impact. There is a clear relation between the probability of budget overruns and the number of minor bidding events during the budget period. This result is expected since it is highly likely that the minor bidding events are triggered by the budget overrun itself.

'Bidding Events' is a reaction to the budget overrun; therefore, the parameter cannot be used to estimate the probability of the occurrence at the time of budget creation. The characterization of the secondary bid events could lead to other managerial implications and should be further analyzed in the future with additional data. As seen in Figure 15, the model shows a correlation between budget overruns and the quantity of bidding events. For example, a lane with only one bidding event has almost 75 percent probability of being under or on budget versus 25 percent for a lane with four bidding events.

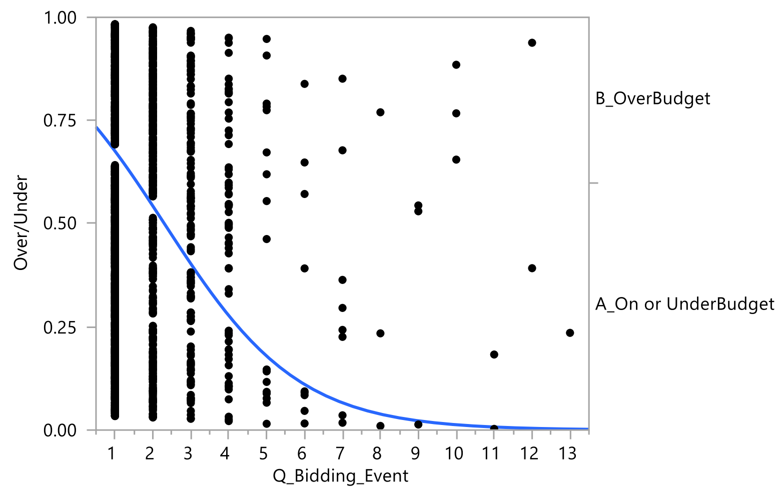

Figure 15: Quantity of Bidding Events vs. Probabilities of Budget Overrun

5. Conclusions

5.1. Summary

The main objective of the project is to characterize the transportation budget failures in terms of significant overruns and understand the reasons for such failures. A model that could accurately predict how close to budget a particular plan will be at the onset could help companies make better transportation budget-based decisions.

Our analysis, even if limited by the data available, shows some clear correlations. As discussed in Section 4, the market conditions at the time of the budget creation are the most significant, followed by the volume variability of the lane. Both parameters are predictors of the future probabilities of budget overruns. Also, some origin and destination states have statistical significance (see section 4), but further analysis with more data is needed to obtain conclusive results.

The main limitations of this research project are related to the lack of precise budget data to build the models on. Even if transportation budgets are commonly used by the companies included in the database, the data is not consistently available. Therefore, we had to rely on a reduced dataset and assumptions. The analysis showed interesting initial results that could be further developed in future research with more budget data.

5.2. Managerial Implications

As a general recommendation, the budget process should contemplate external factors that have a relevant impact on its reliability. The level of accuracy of the budget should be measured systematically. The process should also include a subprocess of performance feedback. This feedback could be used by the shippers to improve the budgeting process.

As a first step, a process or system should be implemented to capture and store the budget data and its performance. This data, together with market and other potential external parameters, could be used to build more precise models in the future and identify additional correlated parameters to monitor. We identify two main parameters for management to focus on to improve the reliability of the budgets: market condition and volume consistency.

Market Condition

Shippers can actively improve the performance of their budgets by considering the market conditions at the time of the budget event. A tight market condition level should be used as a warning that a more precise analysis of the budget assumptions is needed. A soft market condition should be used as a signal for shippers to adopt a more aggressive approach to identify saving opportunities. Being under or on a budget is a positive outcome only if the budget is developed considering realistic and challenging goals.

Tight Market	• The probabilities of market overruns are remarkably high in this scenario • Additional precautions and probably some safety margin should be built in the budgets. • Recommended to identify other factors specific to the lane that might further increase the overrun risk, such as volume consistency.
Soft Market	• The probability of being on or under budget is remarkably high. • This result might be an indication of a too cautious approach to the budget rate calculations • Recommended that shippers adopt a more aggressive approach to commit to identifying savings opportunities. • Being under or on a budget is a positive outcome only if the budget is developed considering realistic and challenging goals.

Volume Consistency

Lanes with inconsistent volume should be analyzed with a higher degree of attention if the yearly volume of the lane is significant. This result is even more significant, considering that we only included the top lanes concentrating 80 percent of the shipper volume. Irregular volume might be influencing carrier asset allocation to the lane, ultimately impacting the budget performance. In case of anticipated irregular volume conditions for a lane, shippers should look for ways to reduce volatility through volume consolidation, better planning, or more accurate forecasting. A consistent, regular volume could be used as an indicator of potential opportunities for more aggressive budget settings.

5.3. Limitations

The analysis is limited by the availability of precise transportation budget data. Although we had an extensive database with more than four years of shipment operations, we did not have the same coverage for budget data. Only a minor part of the records contained budget information in the database.

With additional budget data from additional shippers, we could extend the analysis of the origin and destination states' impact, perform a new test of the economic sector influence hypothesis, and in general, build a model with a higher level of explanatory value.

The available budget information also had some weaknesses since it is geared to regular performance monitoring and not towards formal budget control.

We find evidence that the budget prices were updated several times during the budget period. This behavior is in line with operational performance management and not in line with yearly budget reviews. Some assumptions had to be made on which version of the budget data to use. We had to purposely decide what version of the prices to use as budget values, and we ignored the subsequent changes to the rates.

Due to the nature of the transactional database, we also did not have volume estimations at the time of the budget event. Common industry practice is to use the last-period volume as the budget for the next period. Still, it is also usually adjusted by some factor to reflect anticipated changes in the market or the company's plan for the lane. Also, for some specific lanes, it is highly likely that the shippers use a more sophisticated methodology to analyze the volumes. In some cases, shippers have a formal sales and operations planning process in place with more accurate volume estimations. Due to the lack of volume data, we are not able to identify events of budget overruns caused by relevant changes in the volume of the lane versus the original plan (unexpected high volume or lower than planned). Because of these limitations, we disregarded any in-depth analysis of the volume impact.

We did not have access to the full budget documents; therefore, lanes that are not transacted during the period are not included in the analysis. This limitation eliminated the visibility of the origin/destination lanes that are budgeted to have specified volume and contracted rates but did not materialize (also known as “ghost freight”). These lanes would have been included in the budget and probably in the related bidding events, so it would be interesting to analyze the impact in the future.

We also did not have the precise dates of the budgeting events and had to extrapolate them from the transactional database. It is highly likely that transportation budgets are developed and issued at specific times of the year with a fixed validity period (usually 12, and sometimes 6 months). Relying on the main bidding event as a proxy for the budget event links the start and end dates of the budget to other parameters that might be influencing the launch of the major bidding event itself (such as market conditions or operative needs and issues). This limitation led to having budget periods of different lengths in variable time frames. Even though we consider that the methodology we applied is sound, having precise bidding event dates and scope, and specific methodology would significantly enhance the analysis results.

5.4. Future Research

Future research could address the data shortfalls that we faced during the analysis. More data and more precise information about the budgets would help strengthen the current model and potentially identify additional significant variables.

The preliminary indication of the impact of origin and destination on the budget overrun probability could also be further explored to incorporate regional effects. This research project found some clear correlations between market conditions, coefficient of variation, and budget performance. Even though the analysis is limited by the data available, the insights identified are robust indicators for managerial intervention and future research. The sponsor company can use the project insights to improve its budgeting process. It can also use them to help its customers improve their budgets, which ultimately are used to evaluate the sponsor's own performance.

6. References

Abramson, M. & Sawant, A. (2012). Impact of risk sharing on competitive bidding in truckload transportation (Thesis, Massachusetts Institute of Technology)

Acocella, A, Caplice, C, Sheffi, Y. (2020), Elephants or Goldfish ?: An Empirical Analysis of Carrier Reciprocity in Dynamic Freight Markets

Aemireddy, Yuan (2019, Root Cause Analysis and Impact Of Unplanned Procurement On Truckload Transportation Costs (Capstone, Massachusetts Institute of Technology)

American Trucking Associations (www.trucking.org), ATA American Trucking Trends (2019) Annual Data Compendium

Bleggi, C. & Zhou, F. (2017). A study of freight performance and carrier strategy (Thesis). Massachusetts Institute of Technology.

Bureau of Transportation Statistics (2018), Transportation Statistics Annual Report 2018

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. G. (1996). An optimization based bidding process: a new framework for shipper-carrier relationships (Thesis). Massachusetts Institute of Technology.

Caplice, C. (2009). Electronic Markets for Truckload Transportation. Production and Operations Management, 16(4), 423–436.

Caplice, C, Sheffi, Y. (2003). Optimization-based Procurement for Transportation Services. Journal of Business Logistics, Vol 24, No. 2.

Caplice, Chris. (2012). ESD.S41 Freight Transportation System Analysis [Class Notes]

CH Robinson (2019), US, North American Surface Transportation Market Update Q3 2019

Collins, J. M. & Quinlan, R. R. (2010). The impact of bidding aggregation levels on truckload rates (Thesis, Massachusetts Institute of Technology)

Council of Supply Chain Management Professionals (CSCMP)(2019). State of Logistics Report

Kim, Y. J. (2013). Analysis of truckload prices and rejection rates (Thesis, Massachusetts Institute of Technology)

Pickett, Chris. Navigating The US Truckload Capacity Cycle: Where Are Freight Rates Headed And Why? Journal of Supply Chain Management, Logistics and Procurement 1(1):1-18, 2018

Shmueli, G., Patel, N. R., Bruce, P. C., & Torgo, L. (2017). Data Mining for Business Intelligence: Concepts, Techniques, and Applications in R

Appendix A: Data Dictionary of Loads Data Set

Number	Column Name	Description
1	LoadNum	Unique ID on load
2	BranchCode	Customer Master ID. Shipper.
3	CustomerCode	Customer ID. Depends on Branch Code. One company may have different customer codes based on the division.
4	Vertical	Customer Industry
5	OriginCode	Warehouse code of origin
6	OriginCity	Origin City
7	OriginState	Origin State
8	OriginZip	Origin Zipcode
9	DestinationCode	Warehouse code of destination
10	DestinationCity	Destination City
11	DestinationState	Destination State
12	DestinationZip	Destination Zipcode
13	EnteredDate	The time the order is entered into the system
14	ActivityDate	Initially intended pick up date
15	Rate	Cost of load. Rate = Line haul + Detention +Fuel +All Other Accessorial Charges
16	Mode	Transportation mode: Van
17	CarrierCode	Unique T code for carrier
18	Miles	Miles
19	ActualWeight	Actual weight of goods
20	OrdPieces	Pieces on order
21	ActualPieces	Actual pieces on order
22	ActualPallets	Actual pallets on order
23	HazMat	Flag for hazardous materials
24	DropTrailerFlag	Flag for drop trailer
25	Expedited	Flag for expedited load (urgent order)
26	Volume	Load volume by cubic feet
27	LineHaul	Linehaul rate on load
28	Detention	Detention cost on load
29	Fuel	Fuel surcharge
30	Accs	Accessorial charges
31	QuoteID	Quote ID, key to Cost Quotes Data Set
32	Primary	if the carrier is identified as primary (not always used)
33	Rank	if freight doesn't have a ranking number is not awarded
34	SeqNum	It starts with 0 as the first tender. A "perfect" load will show 0 as the tender sequence because the first tender was accepted.
35	LanePercent	% of loads assigned to a carrier
36	Awarded_Freight	if Primary=1 or rank=1 or lanepercent>0 then is awarded else not awarded
37	TenderMethod	if tenderedbyAT=1 then AT if tenderedby=SpotBid then spot bid everything else is Manual
38	Carrier Fleet Type	This tell us the type of carrier
39	Carrier Fleet Size	This tells us the size in terms of tractors for the carrier

Appendix B: Data Dictionary of Tenders Data Set

Number	Header	Description
1	Cost Quote ID	Key to Cost Quotes Data Set
2	Quote Type	Quote Type equal to NULL when manually booked or spot bid booked
3	Effective Date	Effective start date of quote
4	Expiration Date	Effective expiration date of quote
5	Lane ID	This is how the customer defines their lanes in their bid.
6	Budget	Lane Budget Rate
7	Lane Percent	Percentage of lane volume
8	Transite Time	Transite Time
9	Max Loads/Day	Max Loads/Day
10	Max Loads/Week	Max Loads/Week
11	Origin Hierarchy	address to address (specific), 5zip to 5zip, city-state to city-state (60), 3zip to 3zip, state to state and all combinations.
12	Destination Hierarchy	address to address (specific), 5zip to 5zip, city-state to city-state (60), 3zip to 3zip, state to state and all combinations.
13	Miles	Miles on load
14	BookType	Type of booking
15	EquipType	Equipment Type
16	TenderedDate	Date of tender sent
17	TenderedByAT	Tenderer by auto tender
18	TenderedVia	Communication method for tender
19	BookedDate	The date that carrier was booked
20	Accs	not fuel not linehaul
21	Fuel	Fuel surcharge
22	LineHaul	Linehaul rate on load
23	UpdatedDate	The last time the load is active in the system
24	EnteredDate	Similar to activity date
25	CarrierCode	Unique T code for carrier
26	SeqNum	Number of times the load is tendered, shows how deep into the routing guide
27	vertical	Customer Industry
28	branch	Customer Master ID
29	LoadNum	Unique order number
30	Accepted	Boolean expression
31	TenderMethod	Tender Method