The Road Ahead: Leveraging Truckload Trends for Prescriptive LTL Heuristics

by

Bobby Kheny

Master of Science – Civil and Environmental Engineering
University of California, Berkeley, 2016

Bachelor of Science – Civil Engineering
University of California, Berkeley, 2014

and

Claire Urbi

Bachelor of Science – Agricultural and Applied Economics
University of Georgia, 2011

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 2024

Signature of Author: ______________________________
Department of Supply Chain Management
May 10, 2024

Signature of Author: ______________________________
Department of Supply Chain Management
May 10, 2024

Certified by: ______________________________
Dr. David Correll, PhD
Research Scientist and Lecturer
Capstone Advisor

Accepted by: ______________________________
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

The Road Ahead: Leveraging Truckload Trends for Prescriptive LTL Heuristics

by

Bobby Kheny

and

Claire Urbi

Submitted to the Program in Supply Chain Management
on May 10, 2024 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Anticipating fluctuations in Less-than-Truckload (LTL) volume presents challenges for shippers, carriers, and freight brokers alike. This capstone addresses this issue by developing XG Boost and Explainable Boosting Machine predictive models for LTL volume, leveraging insights from the cyclical nature of demand shifts between Truckload and LTL freight. This study identifies key lags that precede changes in LTL volume by analyzing truckload metrics of Route Guide Depth, Load-to-Truck Ratio, and the Purchasing Managers' Index. The best-performing model, Explainable Boosting Machine, showed the specific inflection points in each metric that predicted the expansion of LTL volume between 1 and 5 months ahead. This model had 75% accuracy and improved upon the industry-standard method of predicting expansion by solely looking at PMI. Explanatory analysis of this model offered valuable insights and informed the creation of practical heuristics for freight brokers, enabling them to respond proactively to market fluctuations. This heuristic had an overall accuracy of 65%. By securing contract rates timelier, freight brokers can effectively navigate changing market dynamics, thus enhancing operational flexibility and adaptability within the industry.

Capstone Advisor:

Dr. David Correll, PhD
Research Scientist and Lecturer

ACKNOWLEDGMENTS

We sincerely thank our Capstone Advisor, Dr. David Correll, for his guidance over the past year. His expertise, professionalism, and unwavering support shaped this research and our semester at MIT.

Next, we thank Christina Carroll, Patrick Darby, Mat Leo, and Derek Matson of C.H. Robinson for their time and knowledge in this project. Their collaboration facilitated the research and underscored the study's findings' real-world relevance.

Lastly, we offer special thanks to our writing advisor, Toby Gooley, for her meticulous feedback and suggestions. Her dedication to coaching us throughout this process was essential in transforming our ideas into scholarly work.

I express my deepest gratitude to my family for their support throughout this capstone journey and my education at MIT. I am also immensely grateful to my capstone partner, Claire Urbi, whose collaboration, insights, and constant challenge to “simplify and accelerate” our model helped refine our product to a practical industry use case. – Bobby Kheny

I would like to first thank my Capstone partner, Bobby Kheny, whose dedication and enthusiasm kept this project on track. Second, it would be impossible to give enough thanks to my husband, Neil Folger, who believed in me from the start. I hope he enjoys this Capstone as much as he would have enjoyed the boat we could have bought with my tuition. Now on to new dreams… - Claire Urbi

TABLE OF CONTENTS

1 Introduction

1.1 Motivation

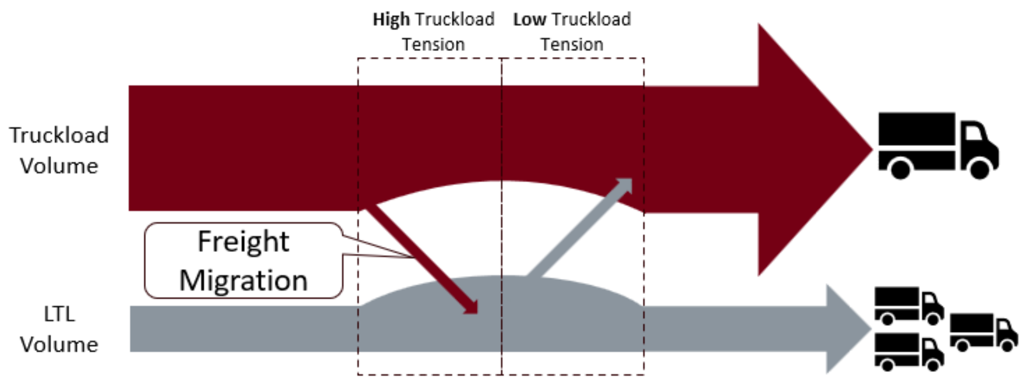
Transportation - the service of moving goods from one location to another - is a critical part of any supply chain. Every step in the physical flow of goods is linked together by transportation. Firms that supply or support the transportation industry work to ensure that all goods, from raw materials to finished products, are moved efficiently to their destination. Trucking is the single largest transportation component, comprising $896B of total annual spending in the United States (Zimmerman et al., 2023). In the trucking market, shippers and carriers seek to balance their speed, cost, and consistency needs.

When transportation demand exceeds capacity, the resulting market tension can push volume from preferred carriers into less desirable carriers or modes. These business cycles impact the shippers' total trucking spend due to increased rates and time to secure transportation. Our sponsor, C.H. Robinson, has a direct stake in building predictions that help anticipate cycles of tension and slack in the Less-than-Truckload (LTL) market. LTL shipments typically consist of one or more pallets and are too large for parcel freight but too small to justify the cost of Full Truckload (TL). LTL carriers can handle smaller quantities of freight in more cost-effective and efficient ways than TL carriers (Moran & Tosi, 2023). Spending on LTL is a $96.3B industry expected to grow from rising demand for smaller shipment sizes (Zimmerman et al., 2023). Industrial production, inventory movement, and e-commerce drive LTL demand (Oglenski et al., 2023). As a broker, C.H. Robinson plays a leading role in the LTL industry by connecting shippers with carriers. They use their expertise to find cost and efficiency savings for LTL freight through value-added services like consolidation and rate negotiation.

The sponsor's motivation for the study is to build off the success of a 2023 Capstone project published by the Massachusetts Institute of Technology's Center for Transportation & Logistics that demonstrated a correlation between TL market tension metrics and LTL volume. The team uncovered a positive, statistically significant correlation for LTL volume with metrics from the truckload contract and spot markets. After evaluating multiple metrics in each category, they identified Route Guide Depth and Load to Truck Ratio as bellwether indicators. Both measures help explain price changes due to supply and demand imbalances. However, Route Guide Depth specifically measures the number of carriers to which a shipper tenders a load before acceptance. Figure 1 visualizes the hypothesis that LTL volume increases as TL shippers seek alternate means of transportation. That capstone also showed a correlation between LTL volumes, tight truck capacity in the TL spot market, and demand. The effect was felt most reliably when volume changes were lagged by 1 to 3 months (Moran & Tosi, 2023).

Figure 1

Hypothesized modal conversion between TL and LTL in the Cyclical Market



The sponsor wants to take further steps toward making market predictions since the correlation between TL market tension metrics and LTL volume has been established. Predicting future LTL volume expansion or contraction would help them advise their customers about future rate increases (Volpe Center, 2018). They could advise their customers to prepare for the change by beginning an LTL Request for Proposal (RFP) process to lock in lower rates ahead of other shippers. Knowing when to begin this process is a competitive advantage, as the sponsor stands to add significant value for their customers. The real opportunity for these shippers is having access to insights that help them integrate transportation as part of their broader strategic procurement goals instead of purely tactical transactions (Caplice, n.d.). Therefore, C.H. Robinson is interested in developing a predictive heuristic – a practical framework – for the modal conversion in demand from TL transportation to LTL.

1.2 Problem Statement & Research Questions

As part of their value proposition, C.H. Robinson is focused on helping their customers achieve financial and strategic goals regarding freight transportation. They already conduct in-depth research and analysis of freight trends, but they want to develop predictive models to improve their customers' experience of the market. Knowing about freight migration from TL to LTL markets can give a shipper time to plan for changing costs and conditions. C.H. Robinson might use the developed framework to enhance current rate or volume models. They might take more active steps with their customers to initiate an RFP process with LTL carriers to lock in lower rates ahead of other firms. Either way, both C.H. Robinson and their customers gain a competitive advantage from insights into the timing of future business cycles.

Gaining those insights will require a predictive model using the established correlations to TL market tension metrics and leveraging our sponsor's data. C.H. Robinson has extensive data due to their history in freight brokerage. This data relates to past events like volumes, rates, capacity, and market samples. When combined with publicly available trucking information, this data can be used to predict future outcomes. We will use these datasets to develop our predictive model and a final heuristic from the model output.

To address these objectives, the questions to be answered include:

- Can a predictive model and heuristic be built to discern future shifts in demand from Truckload (TL) markets to Less-Than-Truckload (LTL) markets based on lagged correlations to TL tension metrics?
- Which TL metrics most strongly influence this prediction, and are there particular magnitudes of TL tension that can serve as a warning indicator for this shift?

1.3 Scope: Project Goals & Expected Outcomes

We hypothesize that an explainable machine learning method such as explainable boosting machines (EBM) will provide a more accurate short-term forecast than traditional forecasting techniques like seasonal autoregressive integrated moving averages (SARIMA) and be better suited to explain its prediction. Additionally, we expect that the lagging correlations discovered by Moran and Tosi in 2023 - Load to Truck Ratio and Route Guide Depth - will contribute the most to the prediction.

This project will build a predictive model for freight migration from TL to LTL markets. This model should be developed to enable informed business decisions, seamlessly integrate into C.H. Robinson's existing business processes, and provide marketable insights to their customers. It should also provide inflection points for TL tension magnitude, where we expect higher certainty of imminent freight migration. We expect this model to assist with negotiating contract rates with LTL carriers and providing better-explained pricing to C.H. Robinson's customers.

Lastly, we will use the predictive model results to create a heuristic that will be provided to C.H. Robinson. To make this heuristic, we will analyze a set of visual metrics in the form of partial dependence plots that explain the inflection points of TL tension that can indicate an imminent demand shift to the LTL market.

2 State of the Practice

This research will require a broad understanding of the trucking industry and predictive modeling techniques. As a starting point, we will review available literature about select topics. Our predictive model must first be made to understand how freight moves through the LTL network. LTL freight's unique path means that not all freight suits this mode. Next, we will explore how the market is cyclical in balancing supply and demand for trucking capacity. As the market rotates through this business cycle, it naturally creates capacity changes that can impact LTL volume and lead to freight migration from other modes. Lastly, we will seek to understand the history and techniques behind time series forecasting. Combined, this knowledge will help us make more realistic predictions about volume migration into the LTL network.

- ***LTL Trucking*** – This section will explore the motivations of LTL carriers during the freight consolidation process and highlight the contrast in freight movement for LTL compared to TL. This review will serve as foundational knowledge for the rest of the project.
- ***Truckload Market Cycles*** – This section will cover the concept of the TL market business cycle and review the factors that drove the most recent business cycle. We will also introduce how TL business cycle data will be used throughout our research.
- ***Freight Migration*** – This section will review shippers' motivations for shifting loads from TL to LTL. Identifying when this shift occurs is one of our sponsor's main motivations and the focus of our predictive modeling efforts.
- ***Time Series Forecasting*** – This section will review the available literature on traditional and machine learning methods of time series forecasting in preparation for implementing those techniques in this research.

2.1 LTL Trucking

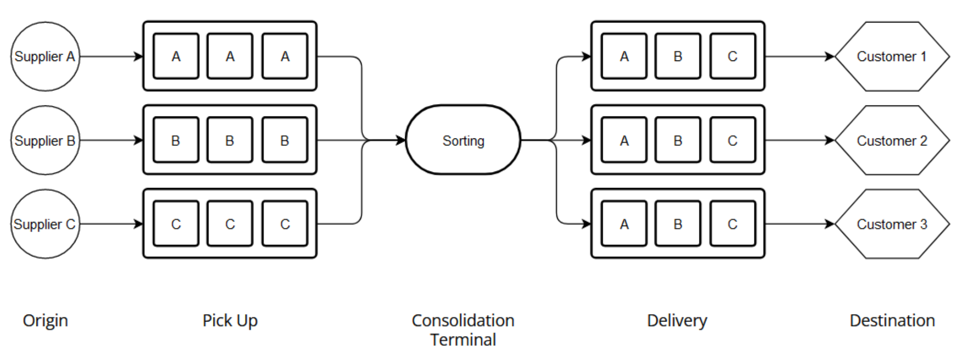
LTL carriers move freight differently than TL carriers. There are more stringent constraints on their ability to flex capacity up and down with demand. Since one of the main subjects of this research is that freight migrates to LTL when TL capacity is tight, we need to explore the LTL trucking process and understand constraints.

LTL carriers can handle smaller quantities of freight in more cost-effective and efficient ways than TL carriers can (Moran & Tosi, 2023). To do this, they need to take advantage of economies of scale to drive down costs and approach the efficiency that naturally occurs with a direct TL shipment. TL carriers employ a point-to-point delivery strategy, which is highly efficient. They pick up a single shipper's load at one origin and drive directly to one destination. LTL carriers do not employ this same delivery strategy because it would leave unused space on each trailer. An LTL shipment typically consists of one or more pallets that are too small to justify the cost of full truckload transport. For any given origin-destination pair, carriers would have the same truck, labor, and fuel costs for a single pallet as with multiple pallets. Trucking is most cost-effective when every available space in the trailer is occupied. Maximizing trailer space allows LTL carriers to increase their number of loads while reducing the average cost per load. They gain efficiency, thus achieving the goal of economies of scale.

To meet this goal, LTL carriers will consolidate shipments. During consolidation, carriers load shipments with similar destinations onto the same truck so they can be transported together. There are multiple ways to structure this consolidation process, but they all involve complexities not found in direct shipments. Both direct and consolidated shipments require loading, unloading, and line haul moves. But compared to direct shipments, consolidated shipments have added complexity from sorting and routing (Caplice & Ponce, 2022). As shown in Figure 2, the most basic form of consolidation has an intermediary stop prior to delivery. A driver will pick up the shipment from the origin and transport it to a nearby terminal, where it will be sorted onto a truck with other shipments on the same delivery route. The complexity of consolidation can increase when a shipment must pass through more than one layer of line hauls, terminals, and trucks before reaching the destination.

Figure 2

Basic LTL Consolidation Process



Note: This figure is adapted from (Fang, 2006).

The consolidation process presents barriers to LTL carriers' ability to respond to changing market conditions. A major barrier is that consolidation requires a strategic investment in infrastructure. In a growing market, increasing capacity might require new terminals or expansion to existing terminals. There is no guarantee that the terminals will be correctly positioned or fully operational in time to meet market needs (Campbell, 1989). The uneven distribution of volume across the network presents another major barrier. Consolidation relies on having enough volume going from each origin to destination to gain economies of scale. An ideal scenario would have frequent and consistent volume across the network. In reality, demand flows unevenly and can fluctuate with shippers' needs. This presents a matching problem between shippers and carriers for shipments between sparsely served areas. To successfully tender loads, in this case, might require higher rates, additional delivery time, or a mix of both.

2.2 Truckload Market Cycles

Moran and Tosi's Capstone (2023) introduced findings that indicated TL market business cycles were correlated to LTL volume. Since this research will build on that original correlation, it is important to look in depth at that business cycle. The TL business cycle, shown in Figure 3, can be described as a constant balancing act between supply and demand. Carriers and shippers are not always able to react instantly to market volatility. Many factors influence the market's ability to gain or shed capacity. The more disruptive those factors are, the longer the resulting periods of imbalance but in general, a single business cycle lasts between 3 to 4 years (Fuller, 2022). Since trucking is an integral part of the U.S. economy, recent business cycles, their catalysts, and the factors that impacted duration are well studied.

Figure 3

U.S. Truckload Market Cycle

Note: This figure is adapted from image by (C.H. Robinson, n.d.).

The most recent business cycle is also one of the most well-studied. We will look in depth at this cycle to highlight real-world examples of forces that impact the transportation industry during each of the four distinct parts. The Covid-19 pandemic disrupted supply chains worldwide starting in early March 2020. Governments, schools, and businesses went into lockdown to slow the spread of the disease. The ensuing drop and then immense surge in consumer demand coupled with labor, trade, and quarantine issues ensnarled the world's supply chains. Covid-19 exposed the deficiencies of the lean inventory models of many modern supply chains because they failed to consider a sustained disruption of this magnitude (Jiang et al., 2021). Lean inventory models, while excellent at driving costs down and efficiency up, leave little in the way of cushion or flexibility for market shocks. Supply chains that employ this model become heavily reliant on each supplier in their upstream network and suffer disruptions if even one of those suppliers fails. Due to these breaks in supply chains worldwide, inventory was backed up, and suppliers looked to the transportation industry to support the flow of goods.

The transportation industry faced immense pressure to provide services throughout the pandemic. The need for increased truck capacity strained carriers. Route Guide Depth and Load to Truck Ratio increased as demand outstripped supply. Objectively, carriers benefitted from high demand through the ability to charge higher prices. Excess demand and a rising percentage of urgent shipments meant that shippers competed on price for available transportation. Working against these positive gains for carriers were chip and parts shortages. Shortages increased lead times on new trucks and drove up prices in the new and used vehicle market. Existing carriers struggled to maintain fleets while potential new carriers were delayed entering the market. Both outcomes delayed the trucking industry's efforts to add necessary capacity. Eventually, rising prices and easing shortages were effective in bringing new entrants to the carrier pool (Croke & Caplice, n.d.).

The trucking market has softened since Q2 2022, when the market became less lucrative for carriers. Excess capacity has dropped the price per mile to nearly break even while carriers' costs increase. Increased costs are due to many distinct factors and can be real or opportunity costs. Factors like the price of diesel, lack of trained truck mechanics, and the deflated value of trucks are all magnifying the effect of low price per mile (Croke & Caplice, n.d.). To end that period of excess capacity, industry experts are preparing for the next equilibrium which would bring the market through a full business cycle since the pandemic began (C.H. Robinson, n.d.).

This most recent business cycle was volatile with pronounced highs and lows. Based on personal communication with our sponsor in 2023, C.H. Robinson felt this tension anecdotally through talking to customers and objectively through gains in LTL shipments. C.H. Robinson characterizes this migration of freight as shipments that would normally move on the TL contract or spot markets being tendered in the LTL market. Our intent is to understand the market trends that led to C.H. Robinson making this observation about freight migration.

2.3 Freight Migration

Freight migration describes a shift in loads from one transportation mode to another. A shift would be impossible to discern at the individual load level among the 53.6M tons of freight moved each day on average in the U.S. (U.S. Department of Transportation, Bureau of Transportation Statistics, n.d.). Migration only becomes noticeable in the aggregate when shippers make similar, sudden, or sustained shifts. This research specifically focuses on the shift of loads moving from TL to LTL. This specific type has not been the focus of much peer-reviewed research, so this section will review more general motivations for shippers choosing LTL over TL. Shippers have individual motivations for choosing one mode over another but can be grouped into proactive or reactive motivations.

Proactive motivations

- ***Price*** - The LTL pricing model of only paying for the space you use can be attractive to many shippers if they cannot consistently make TL quantities.

- ***Urgent delivery needs*** - Shippers may prefer smaller, more frequent shipments based on the urgency of the items. Bottlenecks in supply chains, consumer demand, and type of goods all play a factor in how urgently they need to be transported (Malloy, 2014).
- ***Reduce inventory consolidation for the shipper*** - Inventory consolidation is a method shippers use to delay shipment until they have built up a TL quantity of goods to ship. The tradeoff here is that consolidating inventory leads to increased holding costs. It may be more cost-effective to pay for LTL transportation if the extra holding costs exceed the difference in price compared to TL transport (Fang, 2006).
- ***Specialized services*** - Many LTL carriers offer specialized services to shippers for an additional cost. Shippers may shift to LTL and absorb this extra cost if loads require specialized services. Some examples are specific equipment (liftgate), handling (fragile), or a challenging delivery (congested area).

Reactive motivations

- ***Flexibility in load size*** – When loads do not neatly fit the requirements for TL or LTL transportation, shippers can make judgment calls about mode selection that may not always follow pre-specified rules.
- ***Capacity is not available elsewhere*** – When trucking demand far exceeds supply, shippers will be forced to secure any available transportation, even if that means shipping loads on LTL that are not best suited for that mode (Malloy, 2014).
- ***Accessorial Fees*** – LTL carriers have become more sophisticated in setting fees and are increasingly using accessorial fees to price out shippers with particularly undesirable loads, forcing them into other modes or carriers. An accessorial fee is an extra charge tacked on by the carrier on top of the mileage rate and serves to deter certain types of loads like extreme length loads, for example.

Shippers initiate migration to LTL, but the LTL carriers are not passive in this process. Unlike traditional forms of procurement, the trucking procurement process has the carrier on an equal footing with the shipper. Carriers can accept or reject loads, and shippers may attempt to find a carrier to accept the load more than once. Carriers have leverage when freight is driven to the LTL market by capacity shortages in other modes. They can select the most profitable shipments. Carriers might choose to prioritize who gets capacity based on cost, ease of working with a shipper, or type of freight (Murphy, 2004). As discussed earlier, the capacity for LTL carriers is finite, so these decisions become necessary in a glut of demand.

Through this literature review on market trends, we have established a baseline of understanding to inform our research. We will use this business understanding to make sense of the data from a real-world perspective and ensure that our research findings match reality. Below are specific ways this knowledge will be incorporated into the predictive modeling process:

- ***Data understanding*** – provide context around the shipment-level data from our sponsor.
- ***Data preparation*** – match up known trends to charts of aggregated data to test for validity of preparation steps.
- ***Data modeling*** – make choices around decomposing seasonality and trend components.
- ***Analysis of model output*** – evaluate the resulting inflection points against past business cycles.

2.4 Time Series Forecasting

Since this capstone's methodology will utilize various forecasting techniques, we will explore various time series forecasting models in the following sections. This includes the evolution of time series forecasting from traditional to machine learning and newer techniques such as explainable artificial intelligence (XAI).

2.4.1 Traditional Time Series Forecasting

Many types of traditional models exist to forecast a time series. They range in complexity based on the number of factors used to create the forecast. Basic forecasting techniques developed in the mid-nineteenth century include exponential smoothing, such as Holt's Method, which assumes a linear trend in the forecast (Caplice & Ponce, 2022). In 1970, Box and Jenkins expanded the field of time series forecasting by popularizing the autoregressive integrated moving average (ARIMA) model (Tsay, 2000). And in the modern era, more advanced models exist such as Facebook's Prophet library, which can fit non-linear trends into an additive model (*Prophet*, n.d.).

There are use cases for models at all levels of complexity. Traditional models are quick to set up with results that are easy to interpret. These characteristics can be used to create baseline forecasts whose results can be compared to more complicated model types. If a more complicated model does not improve upon the forecast results of a time series model, it may need to be reworked or reevaluated as the prevailing model altogether (Elsayed et al., 2021). We will explore basic concepts of time series forecasting to understand how the data must be transformed and how we can compare the machine learning model to a baseline.

2.4.1.1 Time Series Patterns and Decomposition

Most time series models take advantage of the data being decomposed into three components. For additive models, this can be depicted in Equation 1, where y_t is the observation (Hyndman & Athanasopoulos, n.d.). The time series components are as follows:

- ***Seasonality*** $(\boldsymbol{S_t})$ – Seasonality is the component of the data that demonstrates a pattern over time. This pattern could occur over a year, quarter, month, week, or other timescale depending on the data, but this timescale must be fixed (Hyndman & Athanasopoulos, n.d.).
- ***Trend-cycle*** $(\boldsymbol{T_t})$ – Trend demonstrates a long-term increase or decrease in the dependent variable (Hyndman & Athanasopoulos, n.d.). Examples of trends include business and economic growth. In contrast to seasonality, a cycle occurs when there are rises and falls in the data that are not over a fixed period (Hyndman & Athanasopoulos, n.d.). An example is the "business cycle," where the economy and industries experience periods of expansion and recession. The motivation for this research—the shift in demand from TL to LTL markets—is cyclical. Although cycles are typically grouped with trends by convention, this does not necessarily need to be true.
- ***Remainder*** $(\boldsymbol{R_t})$ – The remainder is whatever variance is left in the model after removing trend and seasonality. Also referred to as "noise" in the data, this component is stationary and should exhibit no trend or seasonality.

Equation 1

The trend, seasonality, and remainder components in additive decomposition (Hyndman & Athanasopoulos, n.d.).

$$y_t = S_t + T_t + R_t \tag{1}$$

2.4.1.2 Time Series Stationarity and Differencing

As the remainder factor demonstrates, a stationary time series displays no trends or seasonality and no predictable long-term patterns. A non-stationary time series can be transformed into a stationary one by differencing, taking the difference between the current and a past observation, the dataset. A widely used form of differencing is a random walk, which measures the change between each observation in the original time series. Another method is to apply a seasonal difference to the dataset, where each value has the corresponding prior season's value subtracted. Random walk and seasonal differences can both be applied to the same dataset to achieve stationarity (Hyndman & Athanasopoulos, n.d.).

2.4.1.3 Time Series Models

The ARIMA model popularized by Box and Jenkins is one of the most commonly used classes of models in time series forecasting (Hyndman & Athanasopoulos, n.d.). ARIMA models describe the autocorrelations in the data by breaking the calculations into three components:

- ***Autoregression*** – An autoregression model forecasts the dependent variable through a linear combination of past values of that variable (y_{t-i}). Thus, lagged values of y_t are used to infer future values (Hyndman & Athanasopoulos, n.d.).
- ***Moving average*** – The moving average component uses past forecast errors to predict future values (Hyndman & Athanasopoulos, n.d.).
- ***Differencing*** – The final component is differencing the data to transform it into a stationary dataset.

ARIMA can be expanded to account for seasonality by introducing additional terms (Hyndman & Athanasopoulos, n.d.). In such cases, it is called Seasonal Autoregressive Integrated Moving Average (SARIMA) or SARIMAX if exogenous variables are also introduced as independent variables. Since this capstone builds on pre-existing correlations between TL and LTL markets, the TL tension metrics may be considered exogenous variables in a traditional forecasting model.

2.4.2 Time Series Forecasting with Machine Learning

Compared to traditional forecasting techniques, machine learning for time series forecasting is a novel approach with more advanced models continuously published (Masini et al., 2021). Machine learning models like neural networks and XGBoost can outperform traditional forecasts like SARIMA (Agata & Jaya, 2019). Models can factor in more types of information and better establish non-linear dependencies in the data. This performance, however, comes with additional computational requirements and tends to be too complex to explain to the client (Elsayed et al., 2021). Fortunately, increases in information and computational availability have been timely and have made machine learning applications one of the most significant advances for supply chains in the past decade (Mugurusi & Oluka, 2021).

2.4.2.1 Predictive Machine Learning

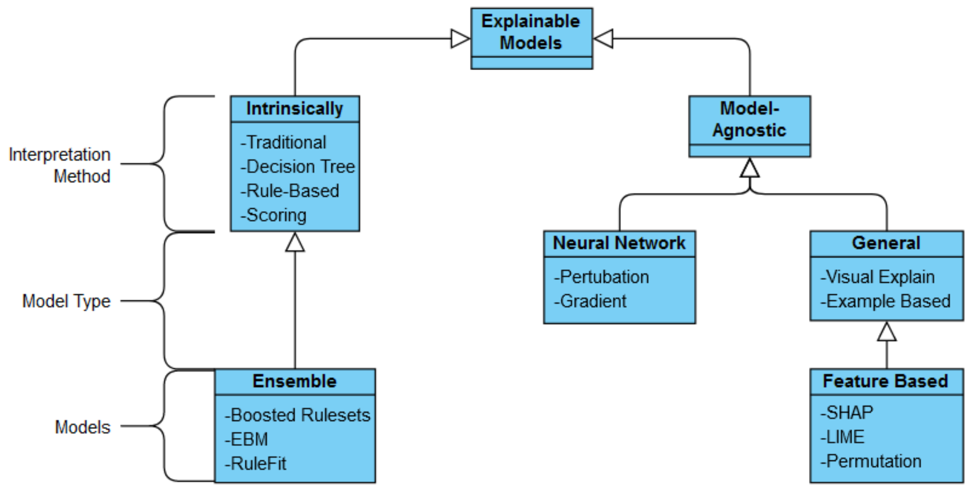
Supervised machine learning uses explanatory variables to predict an output based on relationships present in the data. Several methods have implemented time series prediction, each with its own advantages and limitations for applications. Feed Forward Neural Networks are widely popular and use nodal layers to weigh parameters and fine-tune them to get more accurate predictions. Recurrent Neural Networks can remember the order of input values and assume a dependency on the input sequence (Jaakkola, 2023). Regression trees can predict any unknown function through breakdown classification (Masini et al., 2021). Boosting can be applied to Regression Trees to control for bias and variance in a model (Jain, 2016). In 2021, Elsayed et al. concluded that a conceptually simpler model, such as a Gradient Boosting Regression Tree (GBRT), can sometimes outperform deep neural networks if features and output structures are carefully constructed. Expanding on GBRT, XGBoost makes this approach easier by introducing regularization to the model, reducing overfitting (Jain, 2016). Additionally, the biggest strength of XGBoost is the speed at which the model can be trained (Sheridan et al., 2016).

2.4.2.2 Explainable AI

There is an increasing desire across industries to include explainable AI models (XAI), as it is difficult to trust AI predictions without interpretability and explainability. In supply chain management, decisions must be well understood, as they directly impact the companies' performance. In other words, machine learning models are meaningless unless stakeholders fully understand them (Mugurusi & Oluka, 2021). Societal, rational, and regulatory issues, which are difficult to address by unexplainable machine learning algorithms, can be tackled by XAI (Kamath & Liu, 2021). Figure 4 breaks down a subsection of the existing class diagram for the ever-growing field of XAI models.

Figure 4

XAI class diagram

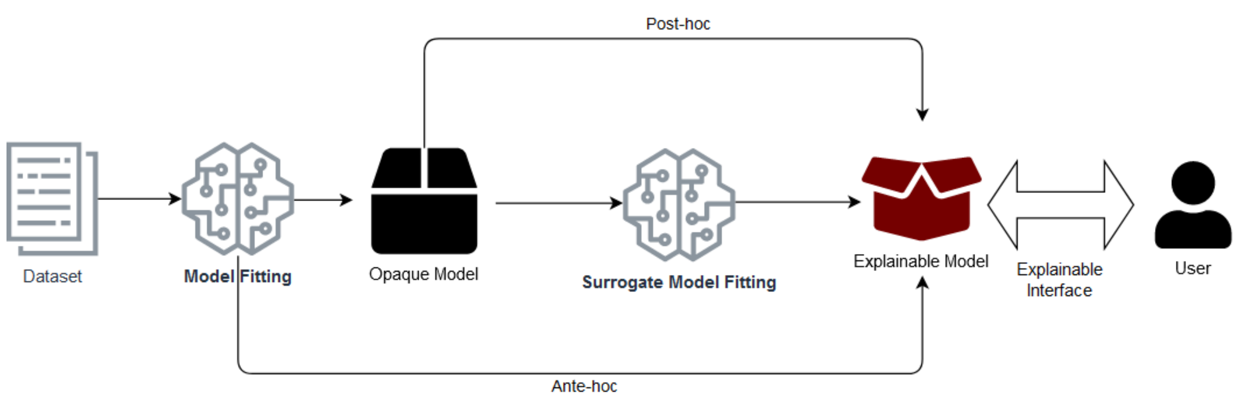


Note: This figure is adapted from (Kamath & Liu, 2021).

Two classifications exist for interpretable models: Ante-hoc models, also known as intrinsically interpreted or glass-box models, natively explain the results. A model that is deemed a 'glass-box' provides not only the forecast output but also an explanation to the end user of what features were the biggest drivers in making up that forecast. Post-hoc algorithms, also called model-agnostic, can be applied to any machine learning model to explain the prediction understandably. As shown in Figure 5, the difference depends on whether the explanation can be generated during the training of the model or the explanation must be derived after model training (Mehta et al., 2023).

Figure 5

Application of ante-hoc and post-hoc XAI



Note: This figure is adapted from (Mehta et al., 2023).

Post-hoc interpretability of models can be generalized as those that provide visual explanations and those that are feature-based (Kamath & Liu, 2021). Visual explanations are a class of models that allow the user to visually inspect how one or many features affect the prediction or other features. A common form of visual explanation is partial dependence (PD) plots, which show how a prediction is influenced by one feature (Relova, 2021). This works by averaging other features based on their marginal distribution. The trade-off with a PD plot is that it assumes the y-axis feature is uncorrelated with the other features (Kamath & Liu, 2021). Features must be carefully selected to avoid co-dependence. Other plots, such as Accumulated Local Effects (ALE) Plots, can address this shortcoming by incorporating conditional distributions (Apley & Zhu, 2019). In the context of this capstone, a visual explanation is useful for demonstrating how a change in TL tension on the x-axis affects the future increase in LTL volume.

Feature-based models quantify the attribution of features to the predicted value and can also consider additional factors, such as the quality of the explanation (Kamath & Liu, 2021). Feature interactions, such as the two-way H statistic, are one way of measuring the effect of feature and model output. They determine the contribution of each factor to the prediction variance. However, they cannot relate the change in prediction to the features (Friedman & Popescu, 2008). Local Interpretable Model-Agnostic Explanations (LIME) explain predictions by approximating the original model locally with an interpretable one (Ribeiro et al., 2016). A shortcoming of LIME is that it may result in a bad approximation and thus create misleading explanations (Kübler, 2023). Another method, Shapley values, distributes feature importance through cooperative game theory. The algorithm's goal is to distribute value fairly based on each feature's contribution to the output (Lundberg, 2018). Yet the downside to Shapley values is that the calculations are computationally intensive and do not explain feature interactions (Kübler, 2023).

In the machine learning community, it is believed that there is a tradeoff between interpretability and accuracy of a machine learning algorithm (Kübler, 2023). However, recent advances in ante-hoc methods have improved accuracy. One such method, explainable boosting machines (EBM), has been shown to compete with or, in some cases, outperform random forest and XG Boost. EBM is a "Generalized Additive Model with automatic interaction detection" (Nori et al., 2019). It separates itself from other generalized additive models by implementing a round-robin training method and a low learning rate. Additionally, EBM can automatically detect pairwise interactions of terms. The developer of EBMs, Microsoft Research, recommends that if one is training a model, they should defer to glass-box models such as EBMs as there is typically little to no loss in accuracy, models can easily be debugged, and explanations are exact and not approximated like post-hoc XAI (Caruana, n.d.). Given this recommendation, this capstone will explore using EBMs to build a predictive model.

3 Data and Methodology

This capstone's scope expands upon Moran and Tosi's findings in 2023. As summarized by Table 1, the methodology for this research will build upon the correlations established by Model 2 of that capstone. Only Model 2 will be expanded into a heuristic to focus this research because it produced the highest correlations. Time did not permit further work on Models 1, 3, and 4, as detailed below.

Table 1

Summary of Scope

	Raw Data	Metrics	Related Models (Moran and Tosi (2023)
Baseline Scope	CHR LTL	Total Shipments or LTL Volume	Model 2 - CHR LTL Metrics Correlation with TL Market Tension at a National Level
	TMC TL RGD	Route Guide Depth (RGD)	
	DAT TL LTR	Load to Truck Ratio (LTR)	
	ISM PMI	Purchasing Managers' Index (PMI)	
Scope Expansion if Time	National LTL	Average Cost (or Revenue) per Hundredweight (CWT)	Model 1 - Public LTL Metrics Correlation with TL Market Tension at a National Level
	DAT TL CPM	Cost per Mile (CPM)	Model 3 - CHR LTL Metrics Correlation with TL Market Tension at a Corridor Level
	CHR Market Tension	Average Weight (or Tonnage) per Shipment	Model 4 - CHR LTL Metrics and TL Market Tension Correlation with PMI at a National Level
		Total Tonnage	
		Average Revenue per Shipment	

The raw data used in this project comes from several different sources. Multiple sources are necessary because using data directly from the source ensures it is not filtered through any intermediary processing. In the case of this project, the data sources are also dictated by the 2023 Capstone. Mirroring the data sources from that research allows this research to draw parallels to the prior correlations. The LTL and TL data come from private sources within divisions of our sponsor company, C.H. Robinson, which we will be redacting for the entirety of this research. The TL Load to Truck Ratio data is publicly available through the DAT Freight & Analytics' (DAT) online platform. This data represents the spot rate market for truckloads aggregated by day. DAT is a U.S.-based transportation information provider that provides fee-based services for carriers, brokers, and shippers. Finally, the Purchasing Manager's Index data comes directly from the Institute for Supply Management (ISM) paid subscription portal. The ISM is a U.S.-based non-profit organization providing education and certification services and industry reporting. In this chapter, we will summarize the included metrics, underlying data preparation steps, and preliminary analysis and then establish how the predictive model is built.

3.1 Metrics

This capstone will analyze several metrics explaining the demand in the TL market and using them as explanatory variables for LTL market predictions of expansion or contraction. While the resulting prediction does not clarify what volume is converting modally from TL to LTL, tightness in the TL market at certain levels will drive shipments to the LTL market to find capacity (Sponsor representative, personal communication, 2023). Moran and Tosi (2023) found that these specific metrics are relevant. However, this analysis will focus heavily on how far in advance these metrics indicate LTL volume growth with the goal of creating a heuristic. The list below represents the metrics that will be explanatory variables for predicting LTL volume.

- ***Load to Truck Ratio (LTR) -*** This is a tightness metric in TL spot markets. A high LTR indicates many more freight requests than available truck capacity. Moran and Tosi (2023) discovered that this metric serves as a bellwether indicator for shifts in TL spot market demand to LTL markets 1–3 months out.
- ***Route Guide Depth (RGD) -*** This metric indicates TL availability for existing contracts. A route guide provides lane-specific options for contracted carriers by order of preference. If the first option is not available, then the second is considered, and so on. A RGD of 1 would mean that the shipper's first choice was always used during the sampled period. A RGD higher than 1 means that the first choice was not always available. For example, a RGD of 1.5 would indicate that the preferred carrier rejected the load 50% of the time. RGD serves as a bellwether indicator for shifts in TL contract market demand to LTL markets 1–3 months out (Moran & Tosi, 2023).
- ***Purchasing Managers' Index (PMI) -*** This index measures the change in economic activity and confidence of purchasing managers in the manufacturing sector based on their monthly survey responses. The trucking industry believes that ISM PMI is a good leading indicator of LTL volume

and thus will be considered in this capstone. Although Moran and Tosi (2023) could not establish a relationship between LTL volume and lagged PMI, we will continue exploring this feature in our predictive model to assess if there is any explanatory value in the index.

3.2 Data Sets and Preparation

The data being analyzed combines C.H. Robinson's data and publicly available information. Table 2 provides a list of the data being analyzed and briefly describes the data preparation steps used for each. Explanations of the three data preparation steps – cleaning, reduction, and transformation, are described further in section 3.2.

Table 2

Detail of Data Preparation Steps

Raw Data	Cleaning	Reduction	Transformation
CHR LTL	One raw data file was produced for each year and was combined to make a single master file	Summed LTL shipments by year-month	Each LTL shipment's data was rescaled on business days for each month. Business days are weekdays, excluding holidays commonly recognized by the transportation industry (*Holiday Schedule | Old Dominion Freight Line*, n.d.).
	One outlier customer was filtered from the dataset, as their rapid growth and decline are not representative of the market		
	Filtered to the LTL outsource customer sample – those customers where CHR facilitates 100% of their shipments		Detrended LTL shipment data
TMC TL	Two tables, TL Loads and TL Tenders, were joined by a unique shipment identifier, Load Number, and from a many-to-one relationship between loads and tenders	The loads were grouped to determine the maximum sequence per load.	Route Guide Depth was aggregated by month
	To include only loads that were successfully executed through the route guide, the table was filtered to include rows with a rate, exclude a spot bid flag, and have a sequence number of less than 20. Limiting to 20 is necessary because some loads will erroneously cycle through the route guide, artificially increasing the spot rate.		
	All sequence numbers were increased by 1 to account for the existing sequence starting at 0.		
DAT TL	A value of 0.00001 was used in place of lanes with zero trucks to avoid dividing by zero	Data was grouped by month	LTR was calculated by dividing the loads per lane by available trucks per lane
ISM PMI	Validated additional years' worth of data and appended to existing data from prior Capstone		

The raw data was cleaned and preprocessed before predictive modeling. Based on research and discussions with the sponsor company, we took as few preparation steps as deemed appropriate. Each step falls into three categories: cleaning, reduction, or transformation.

Cleaning was the first preparation step for all the data sets. Data cleaning is identifying incorrect or inconsistent data within a set and taking steps to correct these issues by replacing, modifying, or deleting data. This is a critical process that must be completed before moving on to any predictive modeling because a model generated on uncleaned data may produce misleading results (Caplice & Ponce, 2022). For this project, cleaning is centered around combining multiple files, joining data sources into a single file, and updating files so they are consistent. Later in the project, several outliers were found to produce significant noise in the dataset, so a second round of cleaning was performed on the data to remove those outliers.

The data was further preprocessed to make it more usable. In the case of this project, the data needed to be reduced through aggregation to make a consistent time series that matched across all four data sets. At the same time, the data was aggregated across customer types so outsource customers – those customers where CHR facilitates 100% of their shipments – could be analyzed against all customers. There was no issue with data size and computational power, a common issue leading to further reduction steps.

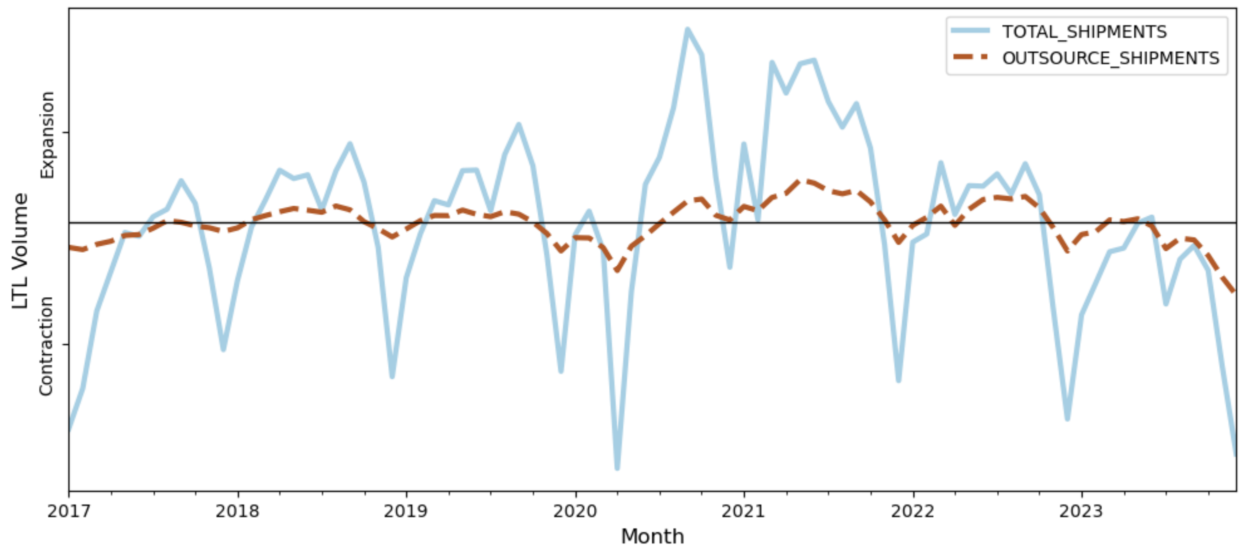
Lastly, the data was transformed into a suitable format for predictive modeling. This project included using existing fields to calculate TL metrics, detrending LTL volume, and scaling the data. Next, we will take a detailed look at the detrending step and then summarize the final data set used in our models.

3.2.1 Detrending

Detrending is a method of data preprocessing that removes overarching trends from time series data so that time series patterns can be identified. The time series data was decomposed to remove trend using the seasonal_decompose function from statsmodels. After the detrending process, the cyclical patterns hover around zero, where positive values indicate expansion of LTL volume and negative values indicate contraction. Seasonality and cyclicality remain in the datasets. Although freight migration is strictly cyclical, the intention is to provide CHR with an early warning system for an increase in LTL volume, which can be affected by seasonality. The detrended values are represented in Figure 6.

Figure 6

Detrended LTL Volume by Customer Sample



Before detrending, CHR's business growth represented a positive linear trend in the time series. We detrended the data to separate the sponsor's normal business growth from the prediction data. Business growth is one possible explanation for overall LTL volume expansion but confounds expansion caused by modal conversion. By removing it, we eliminate that confounding factor to home in on the target variable.

At this point in the research, we continued to run the outsource customer data in parallel with the total customer data, as shown in Figure 5. Moran and Tosi (2023) also explored these two groups and found different levels of correlation to the TL metrics. We aimed to keep both data sets parallel to identify their key differences. As discussed in section 3.3.1, we ultimately focused on outsource customers.

3.2.2 Prediction Dataset Creation

The LTL, LTR, RGD, and PMI datasets are merged on the month_year identifier to create the final dataset used in this analysis. A specific type of merge process, an outer merge, is used to prevent the loss of rows necessary when lagging the features. Each feature had 6 new columns representing lags 1-6 of the metric. For example, LTR shifted by 2 months is represented by the column header "LTR_2" in the dataset. These lags are created to take advantage of the lagged correlations between features and LTL volume, which can be visually observed in Figure 7. Lagged features typically suffer from data loss where the data has been shifted, and no earlier data is available. However, the data in this research had enough earlier data to cover the shift for each metric, resulting in a full monthly time series from 1/1/2017 through 12/31/2023.

Figure 7

Illustration of the original and shifted LTR metric in the resulting dataset

month_year	LTR	LTR_1	LTR_2	LTR_3	LTR_4	LTR_5	LTR_6
1/1/2017	2.251721						
2/1/2017	1.556291	2.251721					
3/1/2017	2.300465	1.556291	2.251721				
4/1/2017	2.377532	2.300465	1.556291	2.251721			
5/1/2017	2.711307	2.377532	2.300465	1.556291	2.251721		
6/1/2017	4.391059	2.711307	2.377532	2.300465	1.556291	2.251721	
7/1/2017	3.303162	4.391059	2.711307	2.377532	2.300465	1.556291	2.251721

We chose to continue using the detrended data for the final prediction dataset. In research where the goal is predicting a set of values, it is useful to use a detrended dataset for analyzing time steps and time series features but then revert to pre-detrended data for the prediction model. In this research, we will develop a heuristic for expansion or contraction instead of a numeric prediction value. Data from the COVID-19 pandemic was not adjusted. We still find a good correlation between the lagged exogenous variables and LTL volume during this period. Finally, we analyzed features from this final dataset to determine how they correlated to each other, as seen in Appendix A.

3.3 Lag Analysis

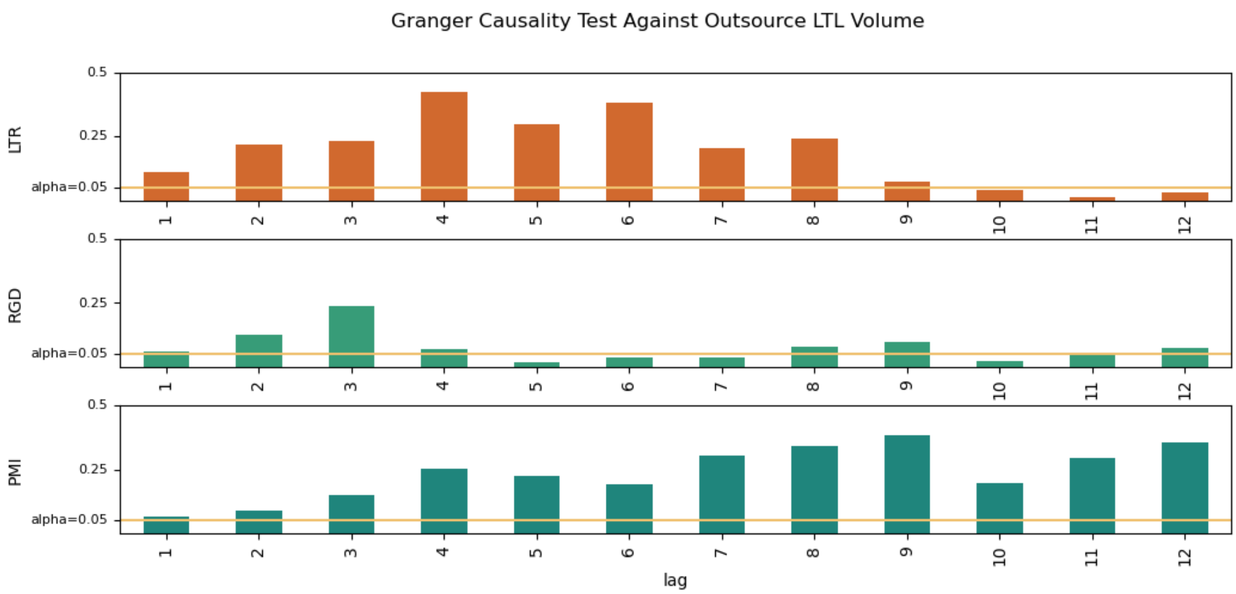
Our analysis of lagged features centers on lags of the features against the LTL volume, which is analyzed with Granger causality tests. We also analyzed the LTL volume's lags against itself to understand the relationship between observations across the time series. The results of this secondary testing can be found in Appendix B.

3.3.1 Granger Causality Tests

Granger Causality is a testing method to determine if one lagged time series statistically correlates with another (Granger, 1969). We ran Granger causality against each feature to determine whether predictive significance exists in the feature and lag. The significant lags have values less than 0.05 (statistically significant). For LTR and RGD, there were multiple options for significant lag, while PMI did not have a lag lower than 0.05. The statistically significant lags for LTR at 10, 11, and 12 months also presented an opportunity to consider other options for this metric. High values for lag risk are not relevant in near-term decision-making and are likely to be leading indicators of seasonality instead of our targeted modal conversion. We did not default to the lag with the best significance but instead used business insight from our sponsor to choose between the top options. The resulting lags were selected at 1 month for LTR and PMI and 5 months for RGD, as represented in Figure 8. This combination of values will be tested in each of our models to give an apples-to-apples comparison of model performance.

Figure 8

Granger Causality Test Against Outsource LTL Volume

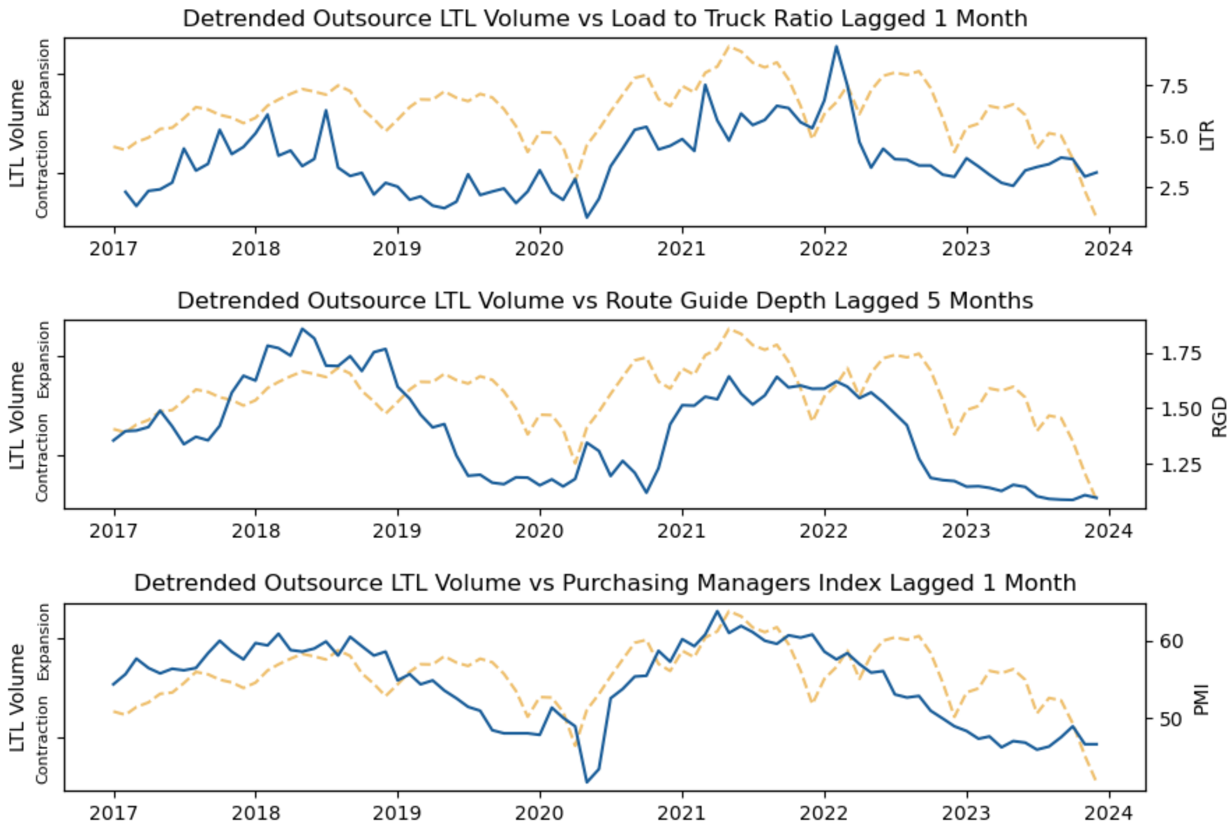


Note: The line represents a significance of 0.05 alpha.

We can plot our target variable time series against these chosen lags for each metric. This gives us a quick visual check to see how well expansion and contraction in the lagged metrics correspond to LTL volume expansion and contraction. The resulting plots are shown in Figure 9.

Figure 9

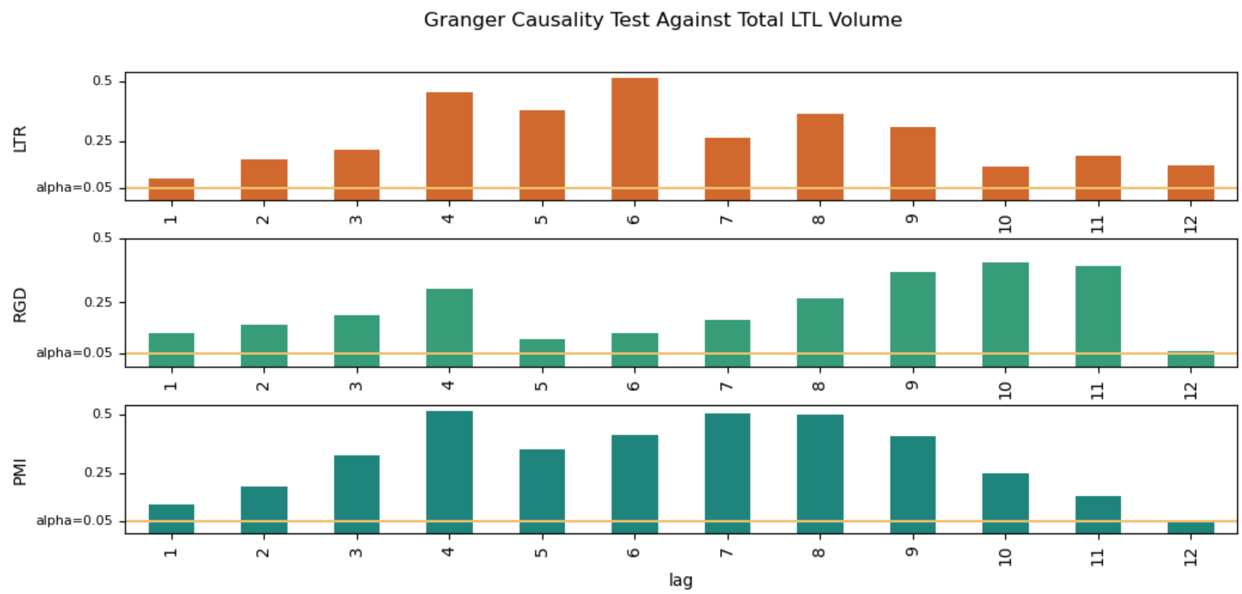
Lagged Features and Detrended Outsource LTL Volume



An important insight from the Granger Causality testing is the lack of a strong correlation between any lags of the three metrics and total customer LTL volume. To recap, we ran the research up until this point on the total data, including all customers, and a subset of that data that only included outsource customers. As shown in Figure 10, no lags meet the statistically significant bar of 0.05 or lower. We discussed this finding with our sponsor to understand the underlying cause. There is a lack of visibility on non-outsourced customer shipments. Non-outsource customers may use the CHR brokerage for any percentage of their LTL shipments less than 100%. That wide range means that even though non-outsource customers make up the bulk of the individual lines of shipment data, we have no visibility into what percentage of those customers are using another broker or the makeup of their transportation needs. Our sponsor indicated that their ability to broker every shipment for a customer allows for faster reactions to market conditions when deciding between modes. This reaction time is a possible underlying cause of stronger correlations to TL market conditions for outsource customers compared to total customers. This insight and discussion led to the decision to continue research efforts on outsource customers exclusively from this point forward.

Figure 10

Granger Causality Test Against Total LTL Volume



3.4 Time Series Forecasting Models

We used three types of models to predict LTL volume: XG Boost Regression (XGB), Explainable Boosting Machines (EBM), and Seasonal Auto-Regressive Integrated Moving Average with Exogenous Factors (SARIMAX). Each model is evaluated with Root Mean Squared Error (RMSE). RMSE is a traditional measure of accuracy that quantifies the average magnitude of prediction error. We take the extra step to convert our results to classifications of “Expansion” – regression value above 0, and “Contraction” – regression value below 0, to establish an accuracy score for each model. This type of scoring means that overfit models are not scored as highly against more generalizable models. This binary accuracy score is the most important measure of our models’ performance. It is critical to prioritize models that correctly predict expansion when LTL volume expands because of the way our sponsor plans to operationalize the final prediction model.

3.4.1 XGB and EBM

We ran XGB and EBM iteratively, testing features lagged from 1 to 6 months and using our hypothesized optimal lag combination of 1 month for LTR and PMI and 5 months for RGD. Given the limited period of data available, each model was cross-validated with only 3 folds. Cross-validation is a technique that runs the model on data that has not been seen before and can indicate how generalizable the model will be. We used cross-validation as an added performance analysis step before layering in the binary accuracy score.

The RMSE for each model and feature combination was compiled for evaluation. We then re-ran each model against a train-test split of January 2017 to December 2021 and January 2022 to December 2023, respectively. This is necessary to evaluate each model's accuracy score.

3.4.2 SARIMAX

SARIMAX is used as a baseline to evaluate the performance of our machine learning models. A machine learning model is expected to outperform SARIMAX models; by comparing the performance, we can judge if our machine learning models are appropriate for this research and if they are running as expected. We ran auto-ARIMA using an ADF test evaluation to determine the optimal seasonal, autoregressive, integrated, and moving average values for both Outsource and Total Volume. We fed in LTR and PMI lagged 1 month, and RGD lagged 5 months as our exogenous variables and then fit the model against the same training set established for our machine learning accuracy score.

3.5 Explainability Analysis

Explainability metrics are devised separately for XGB and EBM since only the EBM model has the explainability of features built into the model from the very beginning. Due to the model type, the baseline SARIMAX model does not have corresponding explainability metrics. We use explainability techniques on the XGB model to understand the decision-making behind the prediction. Explainability metrics seek to show which features are most important to the prediction or how changes in feature value led to different predictions. After our XGB model is trained on optimal features, the model has Shapley values fitted to each feature using the specialized TreeExplainer function in Python. We then fitted partial dependence plots to each feature to evaluate how changes to a feature affect the model's output.

EBM includes explainability features within the package, so this model has no extra analysis steps. EBM's "explain_global" function natively displays metrics similar to Shapley and partial dependence plots through its Dash renderer. Viewing similar explainability metrics for each model allows for side-by-side comparison of the feature metrics.

4 Results

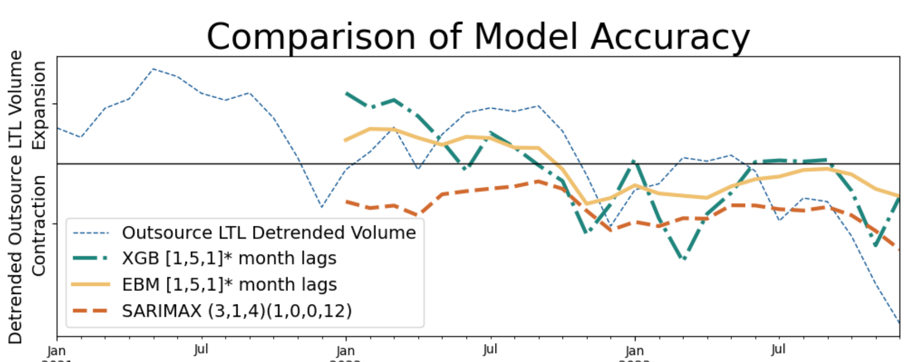
In this section, we present the results of each model's output and the explainability of each machine-learning model. Before running the machine-learning models, we tested them for their generalizability or performance on unseen data through cross-validation. The results of that test are shown in Appendix C.

4.1 Test-Train Split Model Results

After cross-validation, we analyzed the final set of models using the test-train split process. The test-train split process allows the performance of a model to be validated against unseen data. The data set for the target and features is split into a testing set and a validation set. In this research, the data is split sequentially to preserve the time series using 6 years for training and 1 year for testing. A comparison of model performance using the chosen lags (or corresponding parameters for SARIMAX) is visualized in Figure 11. The graph depicts the output prediction from each model compared to the detrended LTL volume in the validation set. We can see that each model performs well at following general trends. SARIMAX and EBM are more conservative predictions, reacting less strongly to peaks and valleys in the cyclical data. XGB is more volatile in predicting volume but more closely matches the magnitude of the peaks and valleys.

Figure 11

Comparison of Model Prediction Accuracy Against Outsource Volume Observations



After reviewing the numerical predictions, we next assess each model's binary accuracy score, which is displayed in Table 3.The binary accuracy measures how often the model correctly predicts expansion or contraction for the following period. For Total Volume, XGB with the optimal lags performs best, with an accuracy score of 88%. For Outsource Volume, EBM, with a lag of 2 months, performs best, with an accuracy score of 79%. Both XGB and EBM significantly outperform SARIMAX in this metric. We will look at explainability metrics for both EBM and XGB, but the EBM model will be used when making our heuristic because of superior accuracy.

Table 3

Binary Accuracy Score of Models

Outsource Volume			
Lag	**XGB**	**EBM**	**SARIMAX**
1	0.63	0.75	
2	0.75	0.79	
3	0.54	0.71	
4	0.63	0.75	
5	0.54	0.83	
6	0.58	0.75	
(1,5,1)	0.58	0.75	0.38

4.2 Explainability

Explainability metrics are devised separately for XGB and EBM since only the EBM model has explainability of features built into it from the very beginning. Explainability metrics for EBM are shown in 4.3.1, while those for the less accurate XGB model are shown in Appendix D. The baseline SARIMAX model is explained through its selection of parameters.

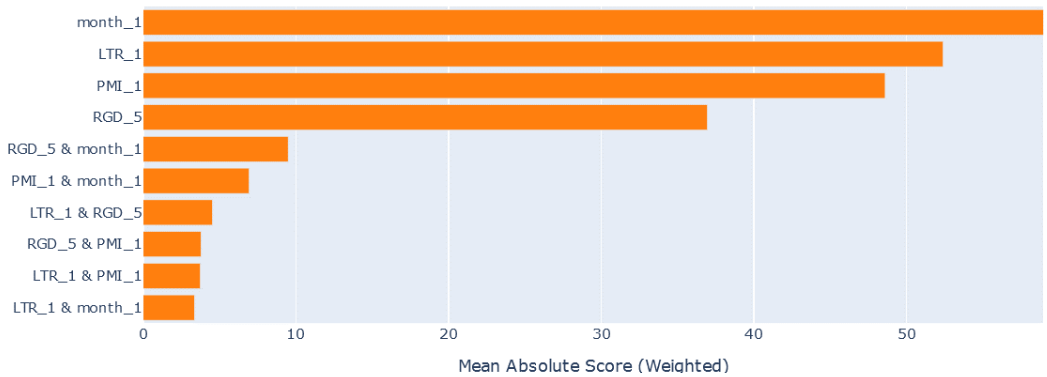
4.2.1 Explainable Boosting Machine

The EBM package output provides each feature's global importance and partial dependence plots. The results for Outsource Volume with LTR and PMI lagged 1 month and RGD lagged 5 months are calculated using this package.

Figure 12 plots global importance and shows month as the most explanatory feature, followed by LTR. These results are dissimilar to the XGB model, likely due to how the decision tree is constructed. However, all features display significance.

Figure 12

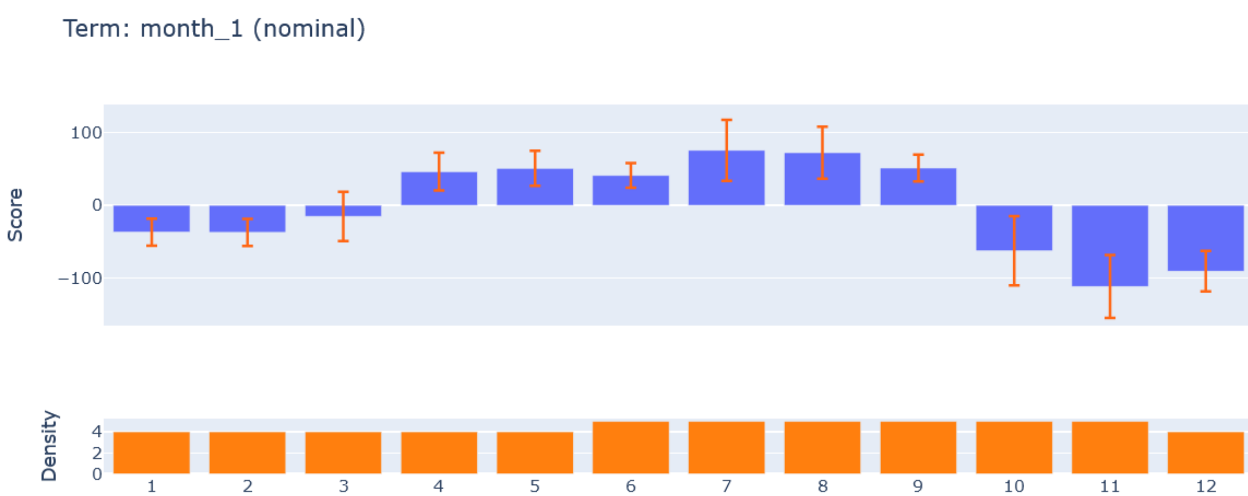
Global Importance of EBM Model



For the partial dependence by month, we observe contraction during winter and expansion during summer, as shown in Figure 13. Although this information could be built into the heuristic, the lagged month says more about the seasonality or cyclicality of LTL volume than it does about modal conversion. The outsize feature importance of seasonality coupled with the existing industry knowledge of season as a predictor of volume both inform our decision to leave it out of the final heuristic.

Figure 13

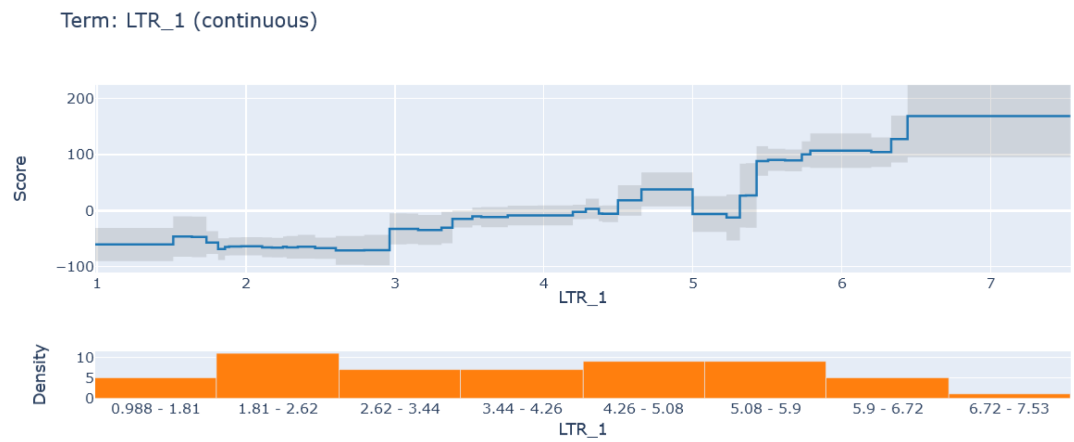
Month Partial Dependence



LTR predicts expansion starting at 4.4, with significant expansion at 5.4, as shown in Figure 14. Contraction becomes significant at an LTR value of 3. According to our heuristic, when LTR hits 4.4, LTL volume may be expected to expand the following month. A user of the heuristic could expect a significant expansion of LTL volume next month if LTR hits 5.4 or higher.

Figure 14

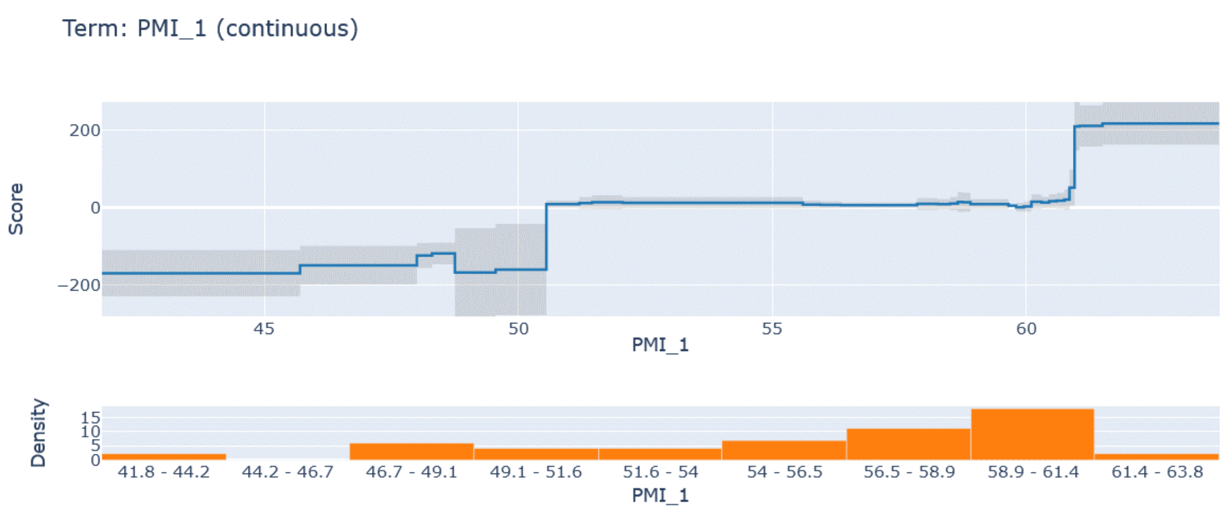
LTR Partial Dependence



As shown in Figure 15, PMI predicts clear expansion at 61 and clear contraction at 51. After discussions with our sponsor, PMI greater than 53 is a more realistic leading indicator of expansion. For our heuristic, this means that when PMI hits 53 or above, expect LTL volume expansion next month.

Figure 15

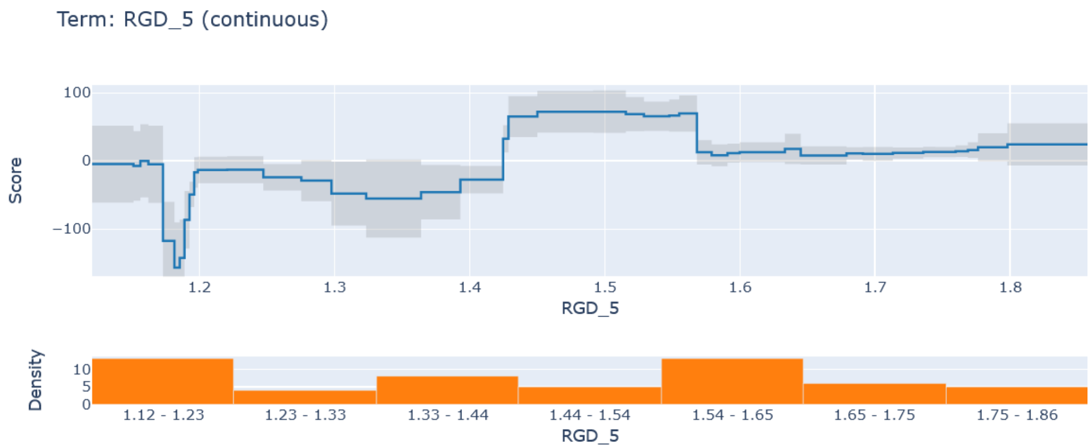
PMI Partial Dependence



RGD predicts a clear divide between expansion and contraction at 1.42, as shown in Figure 16. For our heuristic, this means that LTL volume expansion in 5 months is probable when RGD exceeds 1.42. RGD is ranked lower than the other metrics in terms of feature importance and, therefore, should be evaluated in connection with another metric of greater importance to have the greatest predictive power.

Figure 16

RGD Partial Dependence



5 Recommendations and Discussion

In this section, the final heuristic built from the explainability results in Section 4.2 is introduced. The heuristic, formatted as a matrix, combines inflection points from each of the three TL metrics: Route Guide Depth (RGD), Load-to-Truck Ratio (LTR), and Purchasing Manager's Index (PMI). The inflection points indicate where LTL volume expansion corresponds to changes in the metric value. Section 5.1 will also discuss the binary accuracy score of the heuristic for comparison with predictive models. To conclude the Section, the main limitations of the heuristic are summarized.

5.1 Recommended Heuristic

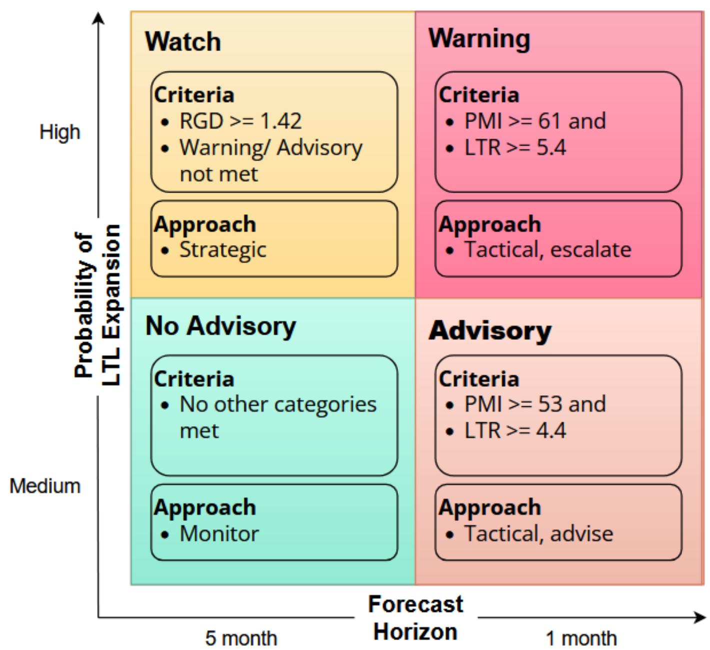
After collecting the inflection points for each of our four features, as presented in the 4.3 Explainability section, we can combine them into a single heuristic. For PMI and LTR, there are two inflection points that represent different levels of confidence. The resulting heuristic separates the inflection points into strategic or tactical recommendations for management, varying by severity. The severity is delineated into Watch, Warning, and Advisory categories similar to those used in predicting future weather events. Categorizing the recommendations in this familiar way helps users understand the level or urgency or seriousness for each scenario. Figure 17 shows the resulting heuristic as a decision matrix.

To use the heuristic:

1. Calculate or collect data points on outsource data subset: RGD, LTR, and PMI
2. Compare the four data points to the indicators in Figure 17.
3. Find the prediction and recommendation in Figure 17 corresponding to the highest severity indicator met by the collected metrics: Warning, Advisory, Watch, or No Advisory.

Figure 17

Heuristic Indicators and Interpretation for Outsource Customers Matrix



Companies using the heuristic matrix can use the forecast horizon to make a strategic or tactical approach when responding to the indicators. RGD, represented by the Watch category, provides an outlook on expansion in the 2 to 5-month range. It is useful for strategic planning, such as advising clients on changes to their network or process improvements. PMI and LTR, represented by the Warning and Advisory categories, provide a short-term 1-month outlook. It indicates when a tactical approach is required. This is the timeframe for shippers to ensure their contract rates are secured.

The process for testing accuracy is to apply the heuristic to the existing data and then calculate binary accuracy against each indicator and prediction of the heuristic. Table 4 shows the calculated binary accuracy. The accuracy for the four predictions is shown without the month component. LTL volume expands and contracts seasonally, so there will be times when it is not appropriate to consider month as part of the heuristic. However, the findings in Figure 20 show strong LTL volume contraction in Quarter 4 of each year and limited contraction in Quarter 1. This seasonal contraction should be assumed to offset any predictions of volume expansion from the heuristic when considering the lag.

Table 4

Calculated Binary Accuracy of Heuristic Indicators and Interpretation for Outsource Customers

	Predicted Behavior				
Actual Behavior	**No Advisory (No Expansion)**	**Watch (Expansion)**	**Advisory (Expansion)**	**Warning (Expansion)**	**Total**
Contraction	20	13	4		37
Expansion	12	24	7	3	46
Total	**32**	**37**	**11**	**3**	**83**
% Observations*	39%	37%	11%	3%	100%
Accuracy**	63%	65%	64%	100%	65%

** Percent of monthly observations that met the indicator criteria from 2017 - 2023*
*** Accuracy calculated as the number of correct predictions of expansion for Watch, Advisory and Warning, and the number of correct predictions of no expansion for No Advisory, over the total number of times the indicator(s) was met for monthly historical data 2017 – 2023*

The heuristic holds an overall accuracy of 65%. Table 5 further summarizes the results of the heuristic as a confusion matrix. Expansion has a precision of 67% and a recall of 74%. No Expansion has a precision of 62% and a recall of 54%.

Table 5

Heuristic Confusion Matrix

		Predicted Label	
		No Expansion	Expansion
Actual Label	No Expansion	20	17
	Expansion	12	34

5.2 Limitations

5.2.1 Modal Conversion as Root Cause is Unproven

One of the stated goals of this project was to predict LTL volume changes as a direct result of modal conversion from the TL market. While the resulting model and heuristic are built on bellwether indicators of TL market capacity issues, this only points to the root cause and does not prove it outright. There are other plausible root causes for the changes in LTL volume we predicted. Despite that, we feel confident that the effect of modal conversion is captured in the models even if other effects are mixed in. Through the pre-processing steps, we worked to limit the possible effects that could compete with modal conversion. Detrending the data was the main way we sought to remove CHR's business growth from the potential list of variables causing LTL volume expansion .

5.2.2 Heuristic for Outsource Customers is a Poor Predictor for Non-Outsource Customers

The original plan for this project was to create a predictive model and heuristic that applied to all CHR customers. During the analysis phase of this research, we noticed that the predictive power of RGD, LTR, and PMI metrics was strongest for the outsource customer subset of data. Based on this finding, we narrowed the scope of the heuristic to the outsource customer subset. This choice reduces the number of customers that can benefit from the heuristic but also allows greater confidence in the resulting prediction for outsource customers.

As shown in Figure 10, the lagged metrics lose significance when run against the data subset of Total Customers' LTL volume. Compared to the same analysis run on outsource customers in Figure 8, there is only one significant lag for PMI_12 and none for RGD and LTR. This result indicated that the predictive model and heuristic using these lagged metrics would have limited predictive power depending on what type of customers were included.

We confirmed the limited prediction power of the three TL metrics on Total LTL Volume by running side-by-side comparison models with outsource customers. Figure 19 shows the output of the three prediction models using the chosen lags. To the eye, these predictions look like those for outsource customers. But the difference is clear when we compare the binary accuracy results in Table 6 from the top-performing EBM model. The EBM models for total customers underperform those for outsource customers subset. The total customer models predict expansion or contraction correctly 67% of the time compared to the same model for the outsource customer subset which predicts it correctly 75% of the time.

Figure 19

Comparison of Model Predictions for Total Customers

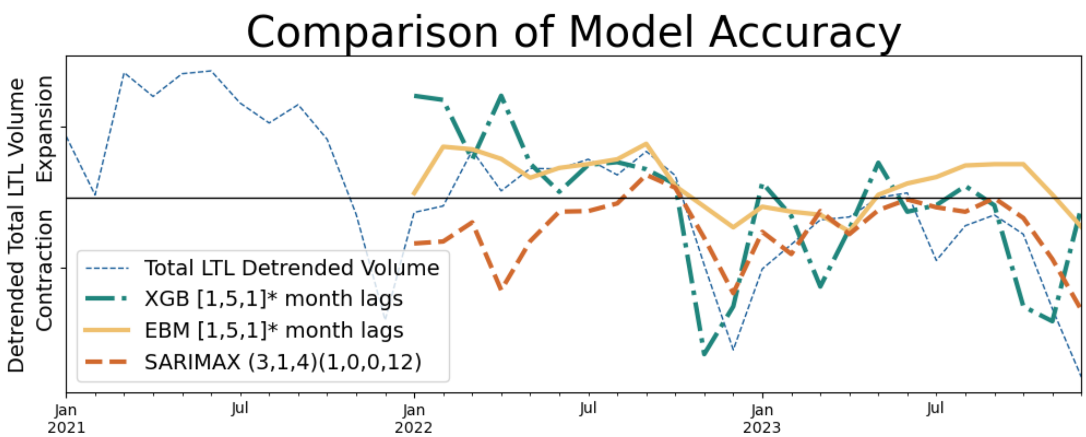


Table 6

Binary Accuracy Score of the EBM Models for Total and Outsource Customer Subset

EBM Model Binary Accuracy		
Lag	**Total Customers**	**Outsource Customers**
1	0.71	0.75
2	0.63	0.79
3	0.54	0.71
4	0.58	0.75
5	0.63	0.83
6	0.79	0.75
(1,5,1)	0.67	0.75

5.2.3 Limited Lag Significance during Market Equilibrium

We found significance for specific lags of each TL metric during our initial analysis. This allowed us to choose the best combination of lags as inputs to the prediction models. That analysis was conducted across the full-time series from January 2017 through December 2023. We were interested in whether those same chosen lags would be generalizable to other periods. To test this question, we re-ran Granger causality during sustained market confidence (PMI value greater than 50). This test corresponded to June 2020 to October 2022. The test output, Figure 19, shows that our lagged features lose significance in this period. The lagged TL metrics are not good bellwether indicators during this period for any of the tested lags (1 through 8 months). Similarly, we tested sustained market uncertainty (PMI less than 50) from November 2022 through December 2023. This test, as shown in Figure 20, showed no significance for LTR and PMI. RGD is significant at lags of 2 and 3 during periods of sustained market uncertainty. This indicates that the predictive model and heuristic can have limited predictive power depending on the status of the market.

Figure 19

Granger Causality During Period of Sustained Market Confidence, Outsource Volume

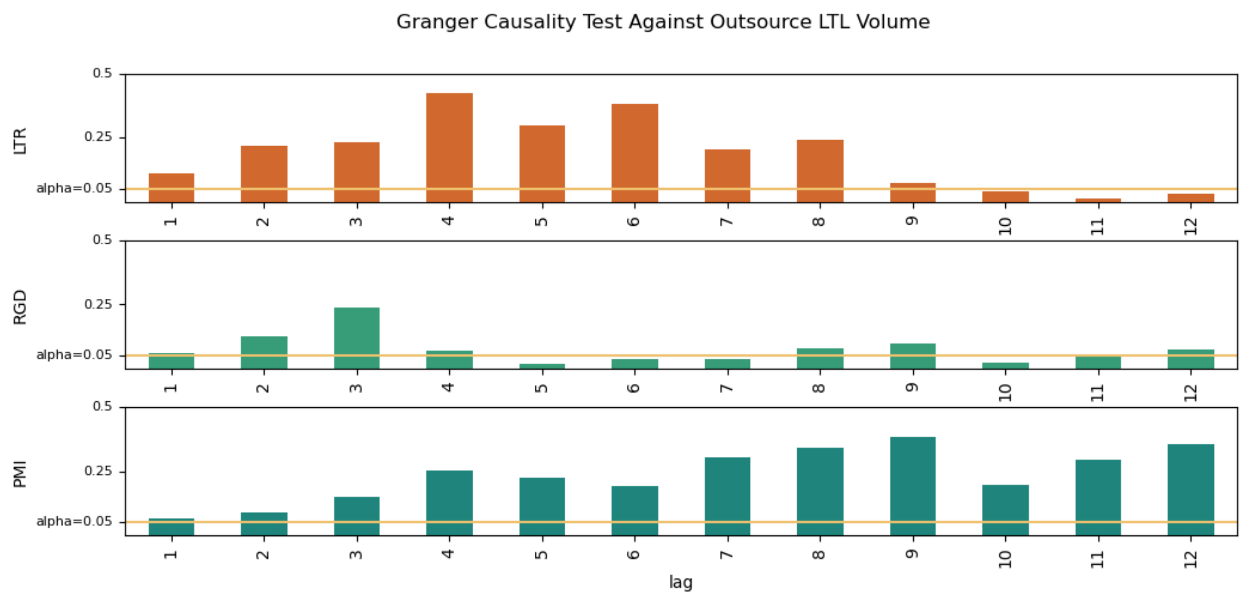
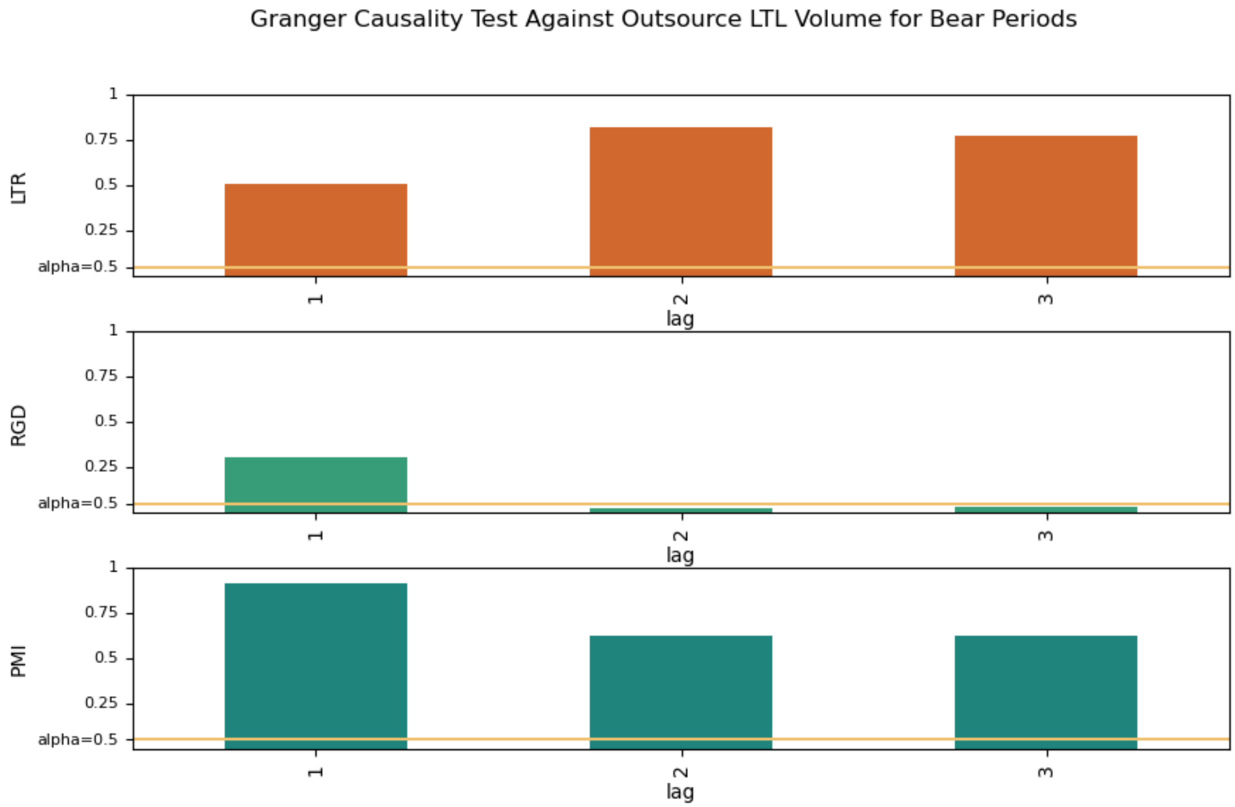


Figure 20

Granger Causality During Period of Sustained Market Uncertainty, Outsource Volume



6 Conclusion

In the trucking industry, capacity shortages can lead to disruption for shippers and carriers and force certain shipments into a more expensive mode. This forced change impacts the shippers' total trucking spend due to increased rates and time to secure transportation. Certain shippers look for leading indicators of changing market cycles to mitigate these risks. A 2023 Capstone found that TL market metrics describing capacity and market conditions could be used as leading indicators due to their correlation with shifts in LTL volume. This capstone sought to build on those correlations by creating a heuristic that combined the accuracy of predictive machine learning models with an accessible format for business stakeholders. Ultimately, the best-performing predictive models showed the specific inflection points in each metric that predicted the expansion of LTL volume between 1 and 5 months ahead. This model had 75% accuracy and improved upon the industry-standard method of predicting expansion by solely looking at PMI. Further analysis of the results showed that the heuristic had an overall accuracy of 65%

In conclusion, this capstone explored descriptive analytics around LTL volume, developed several predictive models informed by these analytics, built explainability metrics to understand how the models arrived at their prediction, and ultimately used the explainability information to inform prescriptive analytics. The heuristic matrix produces predictions more quickly and easily than re-running the predictive models, enabling actionable insights for companies. This application of TL metrics to predict LTL volume is a new area of study. Stakeholders can receive tangible benefits from enhancing their current market research with these findings allowing them more insight into business cycles and allow them to take action ahead of their competitors.

6.1 Future Research

Future research could focus on enhancing the heuristic for outsource customers subset of data. Because this type of modal conversion is a brand-new area of study, there are opportunities to examine the results and build on them. The current models could be expanded to generate a numeric prediction of LTL volume instead of just a heuristic.

Secondly, future research could differentiate LTL volume swings due to modal conversion from root causes. A limitation of this research is that the heuristic does not specify which LTL volume changes are caused by modal conversion. Using the current data sets, future researchers could look for drops in TL volume that correspond well to increases in LTL volume. This would serve as more concrete evidence of modal conversion for this subset of customers. Future groups could perform further analysis of bear markets and bull markets. If there is a shift in shippers' use of these modes correlated to the cycles of the trucking market, this may indicate true modal conversion.

Finally, there are additional opportunities for future research to expand the findings to non-outsource customers. The main limitation of the current heuristic is that it has lower predictive power for this group, which represents a significant number of loads for the sponsor. Future research could start with a survey of transportation professionals for this subset of customers to get insight into their freight and decision-making. Alternatively, future teams could analyze new leading indicators that predict non-outsource volume increases or have a stronger overlap between the two groups compared to the metrics explored in this research.

7 REFERENCES

Agata, R., & Jaya, I. G. N. M. (2019). A comparison of extreme gradient boosting, SARIMA, exponential smoothing, and neural network models for forecasting rainfall data. *Journal of Physics: Conference Series*, *1397*(1). https://doi.org/10.1088/1742-6596/1397/1/012073

Apley, D. W., & Zhu, J. (2019). *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models* (arXiv:1612.08468). arXiv. http://arxiv.org/abs/1612.08468

Campbell, J. F. (1989). *FREIGHT CONSOLIDATION AND ROUTING WITH TRANSPORTATION ECONOMIES OF SCALE* [University of Missouri-St. Louis]. https://mit.primo.exlibrisgroup.com/permalink/01MIT_INST/sjd9fk/cdi_crossref_primary_10_1016_0191_2615_90_90008_M

Caplice, C., & Ponce, E. (2022). *MITx MicroMasters Program in SCM Key Concepts*. MIT Center for Transportation & Logistics.

Caruana, R. (n.d.). *InterpretML: Explainable Boosting Machines (EBMs)*.

C.H. Robinson. (n.d.). *Key Insights for Guiding Your Transportation Strategy*. Retrieved November 22, 2023, from https://www.chrobinson.com/en-au/resources/resource-center/guides/key-insights-guiding-transportation-strategy/

Croke, D., & Caplice, C. (n.d.). *Freightvine* (107). Retrieved November 20, 2023, from https://freightvine.simplecast.com/episodes/dean-croke-state-of-the-market

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). *Do We Really Need Deep Learning Models for Time Series Forecasting?* (arXiv:2101.02118). arXiv. https://doi.org/10.48550/arXiv.2101.02118

Fang, Y. (2006). *Inbound Freight Consolidation for US Manufacturers at China* [Massachusetts Institute of Technology]. https://dspace.mit.edu/bitstream/handle/1721.1/35538/72823216-MIT.pdf?sequence=2&isAllowed=y

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3). https://doi.org/10.1214/07-AOAS148

Fuller, C. (2022, April 13). *What causes the booms and busts of the trucking cycle?* Freight Waves. https://www.freightwaves.com/news/what-causes-the-booms-and-busts-of-the-trucking-cycle

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Holiday Schedule | Old Dominion Freight Line. (n.d.). Retrieved April 2, 2024, from https://www.odfl.com/us/en/resources/holiday-schedule.html

Hyndman, R., & Athanasopoulos, G. (n.d.). *Forecasting: Principles and Practice* (3rd ed.). Retrieved November 11, 2023, from https://otexts.com/fpp3/

Jaakkola, T. (2023, Spring). *Modeling with Machine Learning: RNN (part 2)* [Lecture Notes]. https://courses.edx.org/asset-v1:MITx+6.86x+1T2023+type@asset+block@slides_lecture11.pdf

Jain, A. (2016, March 1). Mastering XGBoost Parameter Tuning: A Complete Guide with Python Codes. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/

Jiang, B., Rigobon, D. E., & Rigobon, R. (2021). *From Just in Time, to Just in Case, to Just in Worst-Case: Simple models of a Global Supply Chain under Uncertain Aggregate Shocks.* https://www-nber-org.libproxy.mit.edu/system/files/working_papers/w29345/w29345.pdf

Kamath, U., & Liu, J. (2021). *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-83356-5

Kübler, D. R. (2023, November 11). *The Explainable Boosting Machine*. Medium. https://towardsdatascience.com/the-explainable-boosting-machine-f24152509ebb

Lundberg, S. (2018). *An introduction to explainable AI with Shapley values—SHAP latest documentation*. https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

Malloy, M. G. (2014, November 17). *Excess Truckload Freight Spilling Into LTL*. Transport Topics. https://www.ttnews.com/articles/excess-truckload-freight-spilling-ltl

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2021). *Machine Learning Advances for Time Series Forecasting* (arXiv:2012.12802). arXiv. http://arxiv.org/abs/2012.12802

Mehta, M., Palade, V., & Chatterjee, I. (Eds.). (2023). *Explainable AI: Foundations, Methodologies and Applications* (Vol. 232). Springer International Publishing. https://doi.org/10.1007/978-3-031-12807-3

Moran, S., & Tosi, N. (2023). *Unraveling the Relation between Trucking Modes: A Correlation Analysis between Less Than Truckload Metrics and Truckload Market Tension*. MIT.

Mugurusi, G., & Oluka, P. N. (2021). Towards Explainable Artificial Intelligence (XAI) in Supply Chain Management: A Typology and Research Agenda. In A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, & D. Romero (Eds.), *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems* (pp. 32–38). Springer International Publishing. https://doi.org/10.1007/978-3-030-85910-7_4

Murphy, J. V. (2004, January 1). *Better Economy, Tighter Capacity May Shift Trucking To Sellers' Market*. Supply Chain Brain. https://www.supplychainbrain.com/articles/1057-better-economy-tighter-capacity-may-shift-trucking-to-sellers-market

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). *InterpretML: A Unified Framework for Machine Learning Interpretability* (arXiv:1909.09223). arXiv. https://doi.org/10.48550/arXiv.1909.09223

Prophet. (n.d.). Prophet. Retrieved November 30, 2023, from http://facebook.github.io/prophet/

Relova, Z. (2021, March 1). *Explain Machine Learning Models: Partial Dependence*. Medium. https://towardsdatascience.com/explain-machine-learning-models-partial-dependence-ce6b9923034f

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier* (arXiv:1602.04938). arXiv. https://doi.org/10.48550/arXiv.1602.04938

Sheridan, R. P., Wang, W. M., Liaw, A., Ma, J., & Gifford, E. M. (2016). Extreme Gradient Boosting as a Method for Quantitative Structure–Activity Relationships. *Journal of Chemical Information and Modeling*, *56*(12), 2353–2360. https://doi.org/10.1021/acs.jcim.6b00591

Tsay, R. S. (2000). Time Series and Forecasting: Brief History and Future Research. *Journal of the American Statistical Association*, *95*(450), 638–643. https://doi.org/10.1080/01621459.2000.10474241

U.S. Department of Transportation, Bureau of Transportation Statistics. (n.d.). *Moving Goods in the United States*. Retrieved November 25, 2023, from https://data.bts.gov/stories/s/Moving-Goods-in-the-United-States/bcyt-rqmu/

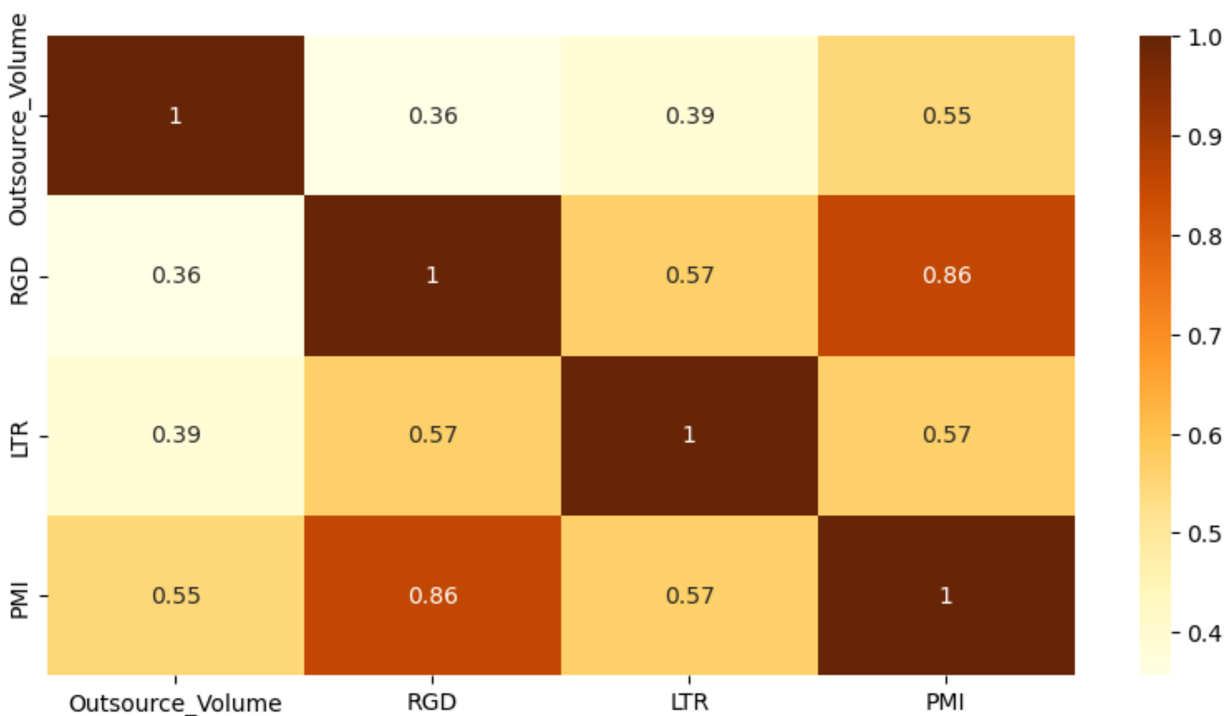
Zimmerman, M., Sonthalia, B., Brogan, J., & Acar, K. (2023). *State of Logistics Report* (p. 63). Kearney.

Appendix A. Multicollinearity Analysis

Multicollinearity is defined as the features of a model being correlated with each other. This can cause issues in models such as linear regression but is not a factor in decision tree models such as XG Boost and EBM. To identify any multicollinearity that might impact our other model type, we ran multicollinearity tests on each feature series using the built-in correlation function of the Pandas library in Python. Figure 21 shows that high multicollinearity exists in the dataset, particularly among PMI and RGD. This indicates that using both features may be redundant in the model. We note this for now and re-evaluate it after observing our explainability metrics.

Figure 21

Correlation Matrix of Features



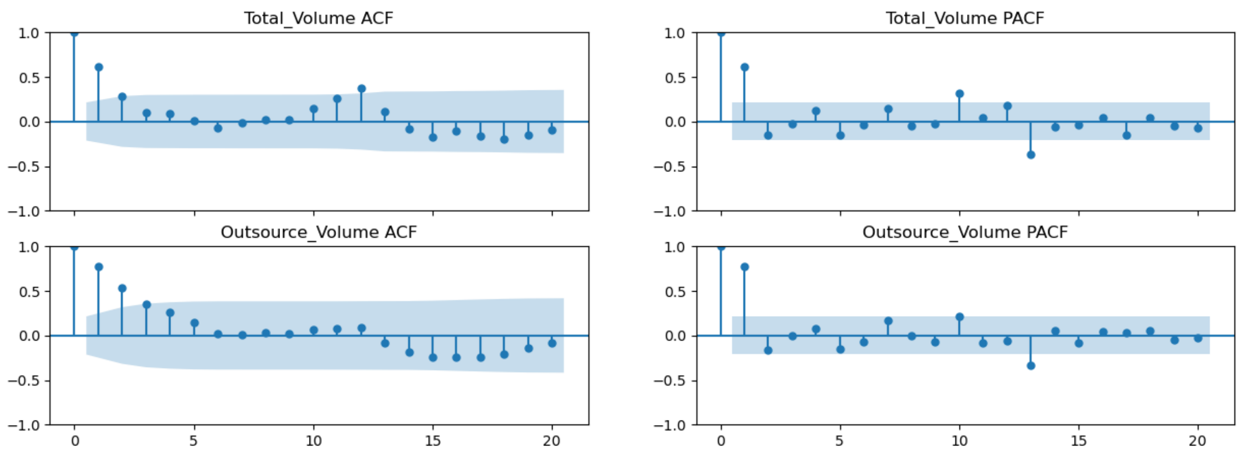
Appendix B. ACF and PACF Plots

We ran the Total Volume and Outsource Volume series through Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots to observe if using our lagged target as a feature in our prediction model was significant. The resulting plots of these two functions against the various possible lags are in Figure 22. ACF shows how lags are related to both the original time series and other lags, while PACF only shows how lags are related to the original time series.

Although lags from 1-2 months proved to be significant, particularly for Outsource Volume, this is excluded from the features in our model since the intention of the model is to discover the predictive strength of the established features.

Figure 22

ACF and PACF Plots of Targets

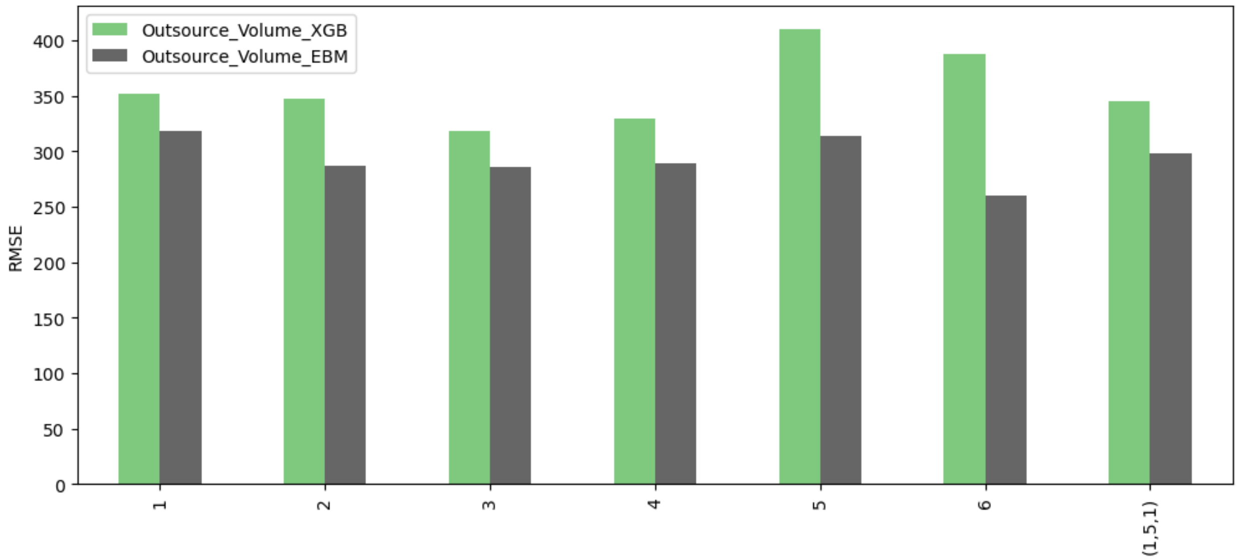


Appendix C. Cross-Validation Model Results

The performance of the cross-validation models is shown below in Figure 23 as RMSE scores. The EBM model has the best RMSE score with a lag of 6 for every feature. Our optimal lags of 1 month for LTR and PMI and 5 months for RGD slightly underperform this measure for the EBM model. EBM outperforms XG Boost for every tested combination. The results of cross-validation models show that we have good prediction accuracy to layer our binary accuracy score in the final test-train split models. The cross-validated RMSE of the baseline SARIMAX model is not shown below because it was not included in the cross-validation process due to the model type.

Figure 23

Three-fold Cross-Validated RMSE by Model Type and Feature Lag, Outsource Volume



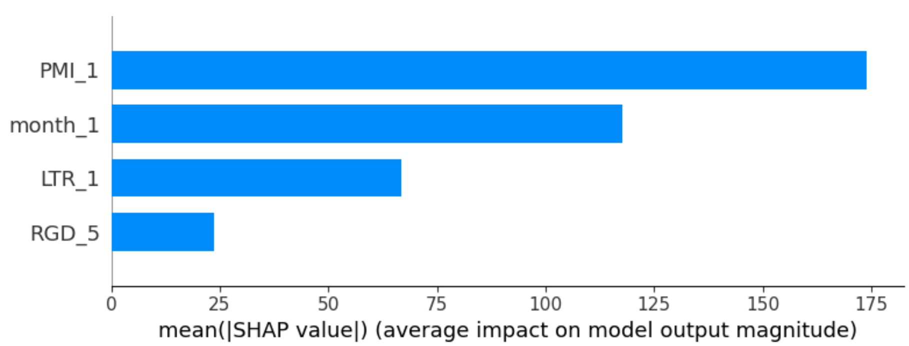
Appendix D. XGB Explainability

XGB Explainability

Figure 24 shows global Shapley values for Outsource Volume with lags of (1,5,1) months. The lags indicate that PMI is the most explanatory feature, followed by month.

Figure 24

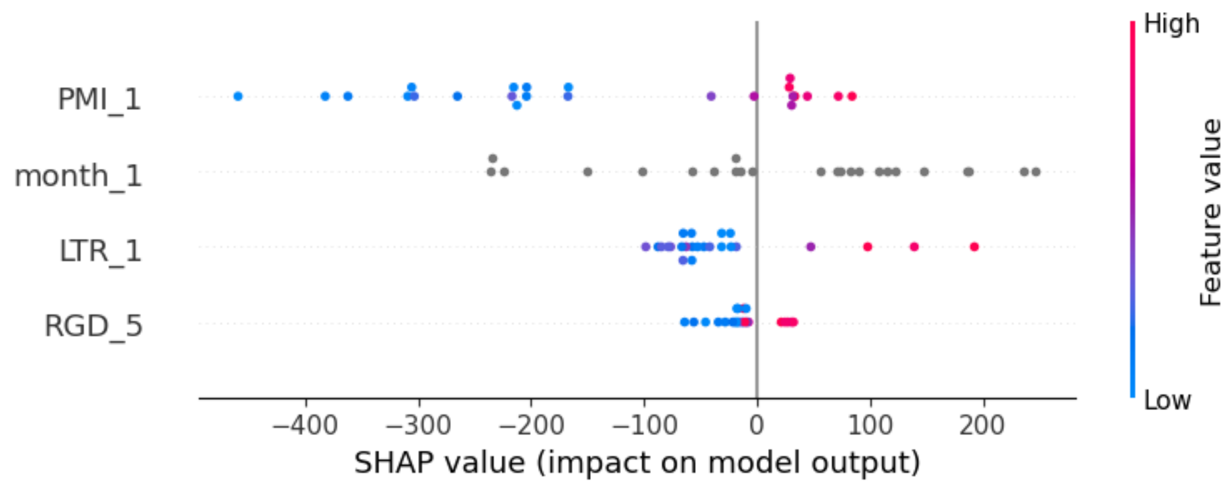
Global Shapley Values for Outsource Volume with Lags of (1,5,1) Months



Diving deeper, we can see those higher values of PMI, LTR, and RGD in Figure 25 correspond with a positive impact on LTL expansion.

Figure 25

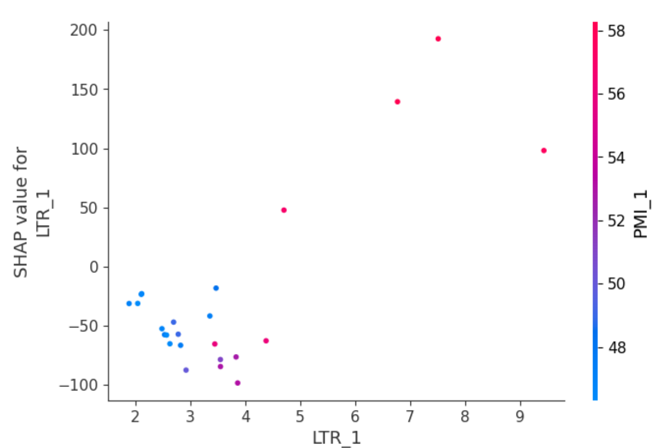
Global Shapley Values with Individual Impact on Model



Correlations in the features are exhibited in Figure 26 by plotting the Shapley dependence plots of LTR and PMI. As LTR increases, both PMI increases and the corresponding Shapley values increase.

Figure 26

Shapley Values Dependence Plot of LTR and PMI



Partial dependence plots in Figure 27 show each feature's impact against the average predicted value. LTR, RGD, and PMI each show a positive trend against LTL, and seasonality is exhibited in the month feature. In particular, the month feature shows contraction in winter months and expansion in summer months.

Figure 27

Partial Dependence Plots of Features