TITLE:

Expanding the freight contract portfolio: Index-based freight contract design under uncertainty

AUTHORS AND AFFILIATIONS:

Angela Acocella, MIT Center for Transportation & Logistics
Chris Caplice, MIT Center for Transportation & Logistics
Yossi Sheffi, MIT Center for Transportation & Logistics

DATE:

March 22, 2022

NUMBER:
Expanding the freight contract portfolio: Index-based freight contract design under uncertainty

AUTHORS AND AFFILIATIONS:

Angela Acocella, MIT Center for Transportation & Logistics
Chris Caplice, MIT Center for Transportation & Logistics
Yossi Sheffi, MIT Center for Transportation & Logistics

ABSTRACT:

Long-term, fixed-price contracts in the truckload transportation sector have been the prevailing relationship form between firms and their for-hire transportation service providers. While advantageous for planning and budgeting purposes, these contracts do not lend themselves well to contexts in which supply, demand, and market uncertainties are prevalent. We propose an alternative, market-based contract for the truckload sector to reduce the unanticipated costs and performance degradation that result from the standard fixed-price contract. With a uniquely detailed and expansive dataset that captures compounding uncertainties, we build empirical transportation provider decision models. We apply them to index-priced demand across a set of design, implementation, and segmentation strategies and present the those that offer Pareto improvements over the status quo fixed-price contract. Finally, we validate our models with a pilot study of indexed contracts implemented by a large agricultural firm in the US. We employ propensity score matching to quantify the causal effect of indexed pricing on transportation suppliers’ contract compliance and costs the firm incurs during a constrained market period.

KEYWORDS:

Truckload freight contracts, index-based pricing, risk-sharing

CORRESPONDING AUTHOR:

Angela Acocella, acocella@mit.edu
Expanding the freight contract portfolio: Index-based freight contract design under uncertainty

Angela Acocella, Chris Caplice, Yossi Sheffi

MIT Center for Transportation & Logistics, 1 Amherst St. Cambridge, MA 02142

Abstract

Long-term, fixed-price contracts in the truckload transportation sector have been the prevailing relationship form between firms and their for-hire transportation service providers. While advantageous for planning and budgeting purposes, these contracts do not lend themselves well to contexts in which supply, demand, and market uncertainties are prevalent. We propose an alternative, market-based contract for the truckload sector to reduce the unanticipated costs and performance degradation that result from the standard fixed-price contract. With a uniquely detailed and expansive dataset that captures compounding uncertainties, we build empirical transportation provider decision models. We apply them to index-priced demand across a set of design, implementation, and segmentation strategies and present the those that offer Pareto improvements over the status quo fixed-price contract. Finally, we validate our models with a pilot study of indexed contracts implemented by a large agricultural firm in the US. We employ propensity score matching to quantify the causal effect of indexed pricing on transportation suppliers’ contract compliance and costs the firm incurs during a constrained market period.

Keywords: Truckload freight contracts, index-based pricing, risk-sharing

1 Introduction

In the United States, firms with goods that need to be moved (shippers) often outsource their over-the-road transportation services to contracted trucking companies (carriers). This outsourced transportation represents a major segment of logistics spend for firms: the for-hire transportation sector totaled $296.1 billion in 2018, or 44.3% of over-the-road freight transportation (AT Kearney, 2018) and represents 43% of total logistics costs in the US (Corridore and Chuah, 2018). The dominant transportation procurement strategy for truckload (TL) shippers and their for-hire carriers results in fixed-price contracts that are non-binding in two ways. First, due to forecasting limitations and fluctuations in end customer demand, shippers are not required to offer or tender their expected volume to their contracted carriers. Second, as carriers must balance their capacity availability and network flows for many shippers’ businesses and effectively face random demand with short lead times (Powell et al., 1988; Berbeglia et al., 2010) they cannot guarantee a truck will be available at the precise time and location for each shipper’s tendered loads.

This is quite unique. In many other sectors, there are often legally enforceable commitments. There are repercussions if a supplier does not provide the amount of the good or service contractually agreed to. Moreover, if the buyer no longer needs the products or services originally contracted for, it is still responsible for paying for all or some of the contracted business.

To mitigate against potential contracted (or “primary”) carrier freight rejections, shippers construct a routing guide for each lane (origin-destination pair) at the time of the strategic procurement event. This ordered list consists of the contracted primary carrier followed by non-contracted backup carriers and is employed at the time of load tender. The shipper first offers the load to the primary carrier. If the primary carrier rejects the load, the shipper sequentially offers the load to each of the backup carriers until one accepts the load. However, the price at which backup carriers
are willing to move the load is usually higher than that of the contracted carrier. On average, shippers end up paying 9-18% more than their contracted price for backup carriers, depending on the external market conditions and how deeply into the routing guide it must go (Aemireddy and Yuan, 2019; Acocella et al., 2020).

Load rejections by primary carriers are exacerbated by the dynamic nature of the TL freight market. The market cycles from periods of over and under supply (see Pickett (2018)) and these overarching market conditions influence how carriers make load acceptance decisions (C.H. Robinson, 2015; Acocella et al., 2020). As shippers’ volumes increase and capacity tightens, contracted carriers’ acceptance ratios (the fraction of freight tendered to them that they actually accept). This is because either the capacity does not exist to serve that demand, or because the carriers can find more attractive, higher priced freight elsewhere.

In addition to the traditional fixed-price contract, shippers and contracted carriers may engage in a portfolio of contractual designs to help either encourage contract stickiness from both parties, or mitigate noncompliance and resulting unanticipated costs described above. In this study, we explore how shippers and carriers can include an alternative contractual form - index-based contracts - into their contract portfolios to help share the risk brought about by shippers’ demand uncertainties, carriers’ capacity uncertainties, and market volatility. A market-based contract has received attention from practitioners. For example, some large players such as Schneider - one of the largest transportation providers in the US - along with others have been pushing to scale these dynamic market-based contracts (Schneider National, 2021). However, its potential benefit has not yet been demonstrated and remaining concerns over how to design such contracts has obstructed widespread implementation.

There has been extensive research on risk-sharing contracts for supply chain partners (e.g., Cachon and Lariviere (2005); Cachon and Netessine (2006); Pasternack (1985)). The combination of demand- and supply-side uncertainties with market uncertainties observed in the TL transportation context, however, has received little attention, despite the sector’s importance to business performance and the overall economy. We aim to address the practical challenges as well as contribute to the literature by demonstrating the contract designs that result in benefits to both buyer and supplier.

The remainder of the paper is organized as follows. In Section 2 we contextualize this work and describe the types of relationships in which shippers and carriers engage. In Section 3 we summarize the relevant literature and in Section 4 we describe our model formulation and the contract policies and designs we test. In Section 5 we detail our carrier acceptance model selection methodology and our steps for choosing the best contract design and policy options. We discuss our modeling results in Section 6. In Section 7, we describe a case study of a large US shipper implementing an indexed contract pilot. We validate our models and conduct a causal inference study on the impacts of these contracts on carrier performance. Finally, we discuss the research implications in Section 8.

2 Dominant shipper-carrier relationships

Shippers typically engage in three main relationship forms with their TL carriers that fall along a spectrum. On one extreme are private fleets and dedicated capacity. These involve the shipper itself owning and running its transportation operations (private) or an arrangement in which an outside transportation provider’s capacity is made exclusively available for that shipper’s needs (dedicated).

On the other side of the spectrum are spot market transactions between shippers and carriers. This is where shippers and carriers interact on a load-by-load basis and the price is determined at
the time of shipment based on prevailing market prices.

Finally, in the middle lie the for-hire contractual relationships. A shipper and a carrier agree to a set price on a lane-by-lane basis established during the shipper’s strategic procurement event. This price applies to all of the shipper’s demand over the course of the contract (typically one year). However, while this standard TL contract is binding in price, it is non-binding in the shippers’ volume commitments and in carriers’ capacity availability. That is, unavoidable demand forecasting errors mean the expected volume may not materialize (referred to as “Ghost Freight”, see working paper by Acocella et al. (2021b)), or more than the expected volume may need to be moved, and the shipper is not required to offer, or tender the expected demand over that time period. At the same time, due to difficulties of ensuring a truck and driver are available at the time and location every load is ready to be moved, the carrier is not required to accept every load it is tendered. Both parties agree to this capacity and volume flexibility because neither wants to be penalized for the uncertainties they face that are largely out of their control.

Within the contracted for-hire relationship, shippers use a portfolio of contract types. The most prevalent is the long-term (1 year or longer) fixed-price contract. However, shippers may use short-term (e.g., 30-90 day) fixed-price contracts in some cases. In addition, shippers and carriers typically engage in more than one relationship form. For example, shippers with their own large private fleets such as PepsiCo and Walmart Stores also supplement their TL demand with for-hire and spot capacity. Alternatively, they may offer their excess capacity to other shippers on “backhaul” lanes in order to balance their network and reposition assets. Further, most shippers engage in both contract and spot relationships. At the same time, carriers may provide services along the whole spectrum as well: dedicated capacity for some customers, contracts for others, and offer capacity on the spot market when they need to reposition trucks to rebalance their networks.

Other contractual forms are in place that help spread the inherent financial risks and reduce uncertainty due to the non-binding nature of the traditional fixed-price contract. Under a tiered volume pricing contract, for example, the shipper and carrier agree to different fixed prices for different volume levels. For carriers, more volume is not always better. A carrier serves many customers; more capacity allocated to one customer, particularly more than what has been strategically planned, throws a carrier’s network out of balance and may cause it to underserve another customer. As economies of scale do not apply, a shipper that offers more volume should not expect to get volume discounts. In fact, shippers may have to pay more for higher volume. Two additional alternative contract forms occasionally used in industry are guaranteed volumes and guaranteed service levels. Under the former contract form, a shipper commits to a set volume tendered to the carrier and if that demand does not materialize, it must still pay the carrier for the unused capacity - either at the load price or some fraction of it. In the latter contractual form, the carrier commits to hauling a set volume. If it cannot take those loads, then it must pay.

In this research, we explore yet another contractual form that aims to mitigate risks and incentivize contract compliance: index-based pricing. Shippers and carriers alike have demonstrated interest in such alternative pricing mechanisms, but few have implemented them. Much of the holdup comes about when discussing the details of how to design the contract.

Thus, we formulate the following research question:

How should shippers and carriers design index-based pricing into their portfolio of transportation contracts that result in a combination of reduced expected costs and higher primary carrier acceptance for the shipper while increasing expected revenue for the carrier?
3 Literature Review

In this section, we discuss the relevant literature on contract and transaction cost economics, supply chain contracting under uncertainty, and freight transportation contracting. The extant literature on contract economics, games, and transaction cost economics is vast, and we do not conduct an exhaustive review. Instead, we highlight the relevant themes and how our freight transportation contracting context fits into the larger bodies of literature.

3.1 Contract theory

In a very general sense, a contract can be described as a game: a contingent plan specifying the set of actions available to each agent and a reward schedule specifying transfers following each sequence of actions in that game. A contract defines the outcome or payment structure of a set of potential future realizations of the world and resulting decisions. If the agents each have private information, or if information arises over time, there may be stages of communication in which information is revealed.

The foundational economic literature on contract theory, which studies how agents construct contractual agreements, typically in the presence of asymmetric information and an array of other uncertainties, is often attributed to Kenneth Arrow. Agents are commonly modeled as maximizers of von Neumann-Morgenstern utility functions – that is, they seek to maximize their expected utility given a set of probabilities of possible future states of the world (von Neumann and Morgenstern, 1947). We draw from this literature by employing an expected cost and revenue framework, describe in Section 4.

3.2 Game theory

The contract literature often seeks the optimal contract in the abstract space of all possible contracts. This is quite a difficult problem. Instead, we take a game theoretic approach, often - though not necessarily explicitly - taken by authors studying supply chain contracting and procurement strategies, in which the performance (e.g., profit or cost outcome) of a set of available contracts and corresponding parameters are compared.

The non-binding TL freight contract is similar to the self-enforcing agreement studied in the context of repeated games literature, which explores agreements that result in both parties maintaining the terms of the agreement over time without the interference of an external party (Telser, 1980). One of the most prominent results of self-enforcing agreements is the Nash equilibrium. A natural extension of this literature considers the conditions under which agreements are vulnerable to renegotiations after repeated rounds of play (Abreu and Pearce, 1991).

3.3 Supply chain contracting and procurement under uncertainty

The theories discussed in the above section have been specifically applied to supply chain procurement. The reader is referred to Cachon and Netessine (2006) for a survey of the game theoretic models applied to supply chain management and a review of models that seek to achieve equilibria, particularly in non-cooperative games.

A common topic in supply chain contract literature is that of channel coordination, which identifies the contracts that result in the players’ Nash equilibrium coinciding with the optimal contract of the supply chain as a whole. Agents seek such risk-sharing contracts to encourage both sides to remain committed to the contract terms. The risk-sharing, or coordinating supply chain contracts that are commonly found in industry and in literature are the buy-back contract and the revenue-sharing contract (Cachon and Lariviere, 2005; Pasternack, 1985). See Simchi-Levi et al. (2014) and Cachon (2003) for review of these models. In the present study, we determine the optimal market-driven index-priced contract design as a means for both shippers and carriers
to share the risks incurred by both sides due to the TL market dynamics and uncertainties. We
then determine whether their use (Pareto) improve, or coordinate, the buyer-supplier channel (as
per Lariviere (1999)).

The benefits of long-term fixed-price versus short-term, dynamic-priced contracts has been
explored in Eppen and Iyer (1997), Wang and Chen (2015), and Cohen and Agrawal (1999). The
authors find that long-term, fixed-price contracts may not always be optimal and discuss conditions
under which short-term contracts may be justified. Specifically, only when market conditions are
stationary - in the mathematical sense in which the mean or variance of the market price do not
change over time - over the duration of the contract terms, then the long-term, fixed-price contract
is best. The non-stationary market condition represents the freight context with which we are
concerned. It underscores why the traditional, fixed-price TL contract may not be suitable given
the dynamic nature of freight markets.

As there are trade-offs between selecting particular contract types, a few authors have consid-
ered a portfolio approach to supply chain contracting. Wang (2011) determines the optimal design
of a portfolio of long- and short-term contracts with access to a spot market from the perspective
of a gasoline supplier. Martínez-de-Albéniz and Simchi-Levi (2005) develops a framework to an-
alyze the performance of a portfolio of supply contracts. The authors construct a portfolio of a
long-term contract, an option contract, and spot market transactions. They design the optimal
contract parameters, and determine the replenishment policy to maximize expected profits in a
multi-period environment. They find that in a general manufacturer-supplier context, the use of
portfolio contracts increase expected profit as compared to traditional single long-term contracts
while also reducing financial risk. However, the authors do not consider the supplier’s revenue
as a result of these contracts. We apply the particular nuances of the TL industry to provide
recommendations for a portfolio contract strategy consisting of indexed and fixed-price contracts.
We further incorporate a channel-coordinating perspective and identify the portfolio of contracts
and designs that Pareto improves shipper and carrier outcomes. That is, either one or both parties
are made better off (e.g., lower expected cost or higher PAR for the shipper and higher expected
revenue for the carrier) without the other being made worse off.

At scale, determining a portfolio contract strategy for shippers is highly complex and may
prove to be analytically intractable (e.g. Cohen and Agrawal (1999)). Thus, we take an empirical
approach to numerically determine best strategies.

3.4 Freight contracts

The uniqueness of freight contracts has drawn attention in the literature. Caplice (1996) makes the
case for the use of contracts between shippers and carriers. The author demonstrates how economies
of scope come about when shippers and carriers make procurement and bid decisions by considering
their network as a whole rather than on an individual lane-by-lane basis. A combinatorial auction
approach is designed to improve efficiencies.

The merits of contracts in trucking are further studied by Hubbard (2001). The author models
the impact of market thickness - that is, number of buyers and sellers utilizing the market - on
shippers’ and carriers’ use of contracts and finds that as the thickness of long-haul trucking markets
increases, simple spot transactions are used by both parties more frequently than are contracts.

Recognizing the uncertainties and price fluctuation in TL transportation and the interest in
alternative, more flexible contractual forms, Tsai et al. (2011) explore the applicability of a futures
or options contracts. The authors find that when demand uncertainty is high, some lanes in the
US lend themselves well to futures contracts by offering value to both shippers and to carriers. The
authors do not, however, explore the characteristics of lanes - or other freight characteristics - that
make market-based contracts most attractive. This latter question is addressed by Acocella et al. (2021a).

Brusset (2009a) and Brusset (2009b) study freight contracting strategies. In Brusset (2009b), given a menu of contract options, the shipper and carrier may choose a single contract to hold over a finite time horizon. Brusset (2009b) includes a form of fixed start-up, or information costs for each contract (i.e., the transaction costs associated with running procurement events and constructing routing guides). Brusset (2009a) discusses the lock-in costs that encourage ongoing interactions between a shipper and carrier when they enter contracts with one another (i.e., the shadow of the future discussed in Acocella et al. (2021a)).

While providing a theoretical foundation for freight contracting, Brusset (2009a) and Brusset (2009b) overlook three critical aspects of the US freight industry landscape. First, the capacity and volume uncertainty introduced by the non-binding nature of freight contracts is omitted. That is, in the author’s formulation, the shipper tenders all realized demand to the contracted carrier and the carrier accepts all of this volume at the contract price. Any demand in excess of the awarded volume is served on the spot market. Of course in reality, shippers may tender above or below awarded volumes and a carrier may reject any amount of tendered volume. We explicitly consider this in our expected cost formulation of the freight contracting problem.

Second, Brusset’s models assume that any deviation from the contract terms results in immediate termination of the contract. In practice, when a primary carrier does reject load tenders, the contract still remains in effect. Moreover, shippers often tender surge volume to contracted carriers. They may pay the same contract rate, or the contracted carrier may accept the excess volume but at a price premium. Alternatively, the excess volume may be rejected by the primary carrier but picked up by a backup carrier.

Finally, in order to avoid trivial solutions in which the shipper always chooses to use the spot market, the author models cost of using the spot market to be higher than that of using any contract. This is not always the case - in particular, during soft market periods. By adopting such an assumption the authors overlook the fluctuations in market conditions that are pervasive in the freight industry (Acocella et al., 2020, 2021a; Pickett, 2018).

In this research, we propose a contracting approach built off of empirical models and captures real-world truckload industry complexities not considered in previous literature. The approach aims to offer practical strategies for shippers under more realistic scenarios representing different suppliers (carriers), freight, and network characteristics.

4 Model Formulation
In this section, we describe the order of the decisions a shipper makes, the contracting design elements and policies we test and how we measure their performance relative to one another.

4.1 Decision sequence
The shipper’s decision as to whether and how to implement an index-based contract on a lane with a carrier is formulated as follows. The shipper’s procurement event corresponds to time $t_0$, when the shipper must decide which one of a set of contract policies, $\pi \in \Pi$, it should implement with a single primary carrier, $C_p$, on a single lane from origin $i$ to destination $j$, that will be in effect for $T$ periods.

In addition to the contract policy and design, the shipper also decides the initialization price, $F_0$, which defines both index-based policies. This initialization price is based off of an exogenously determined benchmark price for the lane: $BM_{i,j,t_0}$. In practice, shippers have access to such benchmark prices through memberships to freight analytics services. It is also information the shipper collects through its reverse auction process. This lane-specific benchmark price completely
defines the Fixed price contract on this lane. We assume the exogenous index to be used is predetermined by this shipper and carrier.

The knowledge or information available to the shipper at $t_0$ is defined by the vector $\kappa$. It contains potential contracted carrier’s service type (i.e., asset or non-asset), the current lane spot price, $S_{i,j,t_0}$, and benchmark price, $BM_{i,j,t_0}$, the current market condition, the previous periods’ tendered volumes, $D_{i,j,t}$, spot and benchmark prices, index values, $I_{t}$, and historical lane characteristics, $L_{i,j,t}$ (as defined in Section 5.2) all for $t < t_0$. The shipper also has an understanding or expectation for each subsequent period of its future demand, spot market and benchmark prices, and the probability the contacted or primary carrier accepts loads at $P_\pi$, the contract price for that load defined by the policy chosen in $t_0$ - that is, $Pr[A_k | C_p, P_\pi]$.

In each subsequent period, $t > t_0$, the index value, spot price, and actual demand are realized. Each load, $k$, is first tendered to the primary carrier, $C_p$, at the contracted price, $P_\pi$. The primary carrier accepts the load with probability $Pr[A_k | C_p, P_\pi]$. As we discuss in Section 5.2, this primary carrier acceptance decision depends on a set of load, lane, and carrier characteristics.

If the primary carrier accepts the load, the shipper pays the offered price, $P_\pi$. If the primary carrier rejects the load, the load is offered to backup carriers, $C_b$. When shippers must rely on backup carriers after primary carrier load rejections, they pay a ‘backup premium’, $bp_t$, or a price escalation above the fixed contract price on the lane. The value that the backup premium takes on may reach up to 18% or more depending on how deep into the routing guide the shipper must go until a backup carrier accepts the load and the current market conditions (denoted by $t$) (Aemireddy et al., 2020; Aemireddy and Yuan, 2019). We model the price a shipper pays backup carriers as $P_{b,t} = bp_t \times BM_{i,j,t_0}$. When tendered a load, $k$, the backup carrier accepts the load with probability $Pr[A_k | C_b, P_{b,t}]$.

If the backup carrier accepts the load, the shipper pays $P_{b,t}$, otherwise the load is offered on the spot market at the current lane-specific spot price, $S_{i,j,t}$ and is accepted with probability 1. The expected cost setup encapsulates the non-binding nature of the shipper-carrier contact - that is, contracted carriers reject loads with some probability. For each load, $k$, and policy, $\pi$, the shipper has expected cost:

$$E[\text{cost} | \pi]_k = Pr[A_k | C_p, P_\pi] \times P_\pi + (1 - Pr[A_k | C_p, P_\pi]) \times Pr[A_k | C_b, P_{b,t}] \times P_{b,t} \quad (1)$$

At the end of the fixed time horizon, $T$, the performance of each contract strategy and contract design can be summarized by two metrics (1) the total expected cost per mile (CPM) of the strategy:

$$E[\text{CPM} | \pi] = \frac{1}{K} \sum_{vk} \frac{E[\text{cost} | \pi]_k}{\text{Dist}_k} \quad (2)$$

where $K$ is the total number of loads realized over $T$ and $\text{Dist}_k$ is the distance of the lane on which load $k$ is tendered, in miles; and (2) the average primary carrier acceptance ratio over the time horizon:

$$PAR_{C_p, \pi} = \frac{1}{K} \sum_{vk} Pr[A_k | C_p, P_\pi] \quad (3)$$

Shippers want low expected costs, and high primary carrier acceptance. Thus, the shipper will choose the Indexed contract that offers PAR and expected cost at least as good as the benchmark
Fixed price contract. On the other hand, the carrier will agree to the Indexed contract if it results in contract price at least as good at the benchmark Fixed price contract. Thus, we present the design parameters values that offer a Pareto improvement over the Fixed price contract.

4.2 Contracting policies

We define the set of contract policies to test based on input from practitioners and previous literature. The indexing policies must be simple enough to implement and agreeable for both shippers and carriers. Our aim is to demonstrate how a small set of indexed pricing policies compare to the baseline, or status quo policy, which is the long-term fixed-price contract. Further extensions of this research may explore additional policies.

Our two policies are defined below and summarized in Table 4.1. First, \( \pi_0 \) is the baseline Fixed-price policy in which all loads are offered at the lane benchmark price, \( BM_{i,j,t_0} \). Second, \( \pi_1 \) is the Indexed policy in which all loads are offered at the indexed price at the time, \( In_t \).

Table 4.1: Contract policy summary

<table>
<thead>
<tr>
<th>Policy</th>
<th>Name</th>
<th>Offered Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_0 )</td>
<td>Fixed-price</td>
<td>( BM_{i,j,t_0} ) ( \forall k )</td>
</tr>
<tr>
<td>( \pi_1 )</td>
<td>Indexed</td>
<td>( In_t ) ( \forall k )</td>
</tr>
</tbody>
</table>

4.3 Indexed contract design

We expect that the performance of each index-based policy depends on certain contract design choices that the shipper makes upfront. Our experience with shippers and carriers suggests that the largest barriers to implementation of index-based contracts in the TL industry relate to the design details we discuss here.

4.3.1 Index choice

First, the shipper and carrier must agree on the index to be used. It must appropriately reflect the dynamics of the freight to which it is applied. Further it must be transparent in the data that it comprises and generally accepted across the industry (Bignell, 2013). Within the TL Dry Van freight industry, a few such indexes exist. These include several DAT National indexes, the Cass National Truckload Linehaul Index, the Stephens Freight Index, and the Morgan Stanley Freight Index.

We assume the the index chosen is the Cass National TL Linehaul Index. It is public and freely available consistently from 2005 (Cass Information Systems, 2020). It is comprised of $2 billion in transactions, 95% from contract transactions and 5% from spot transactions on lanes that span the continental US. The index is a good representation of the general industry, as this aligns with published estimates of the industry-wide ratio of contract to spot transactions; Caplice (2007) estimates 90-95% of the total TL market spend is established through contracts.

Further, the Cass TL Linehaul Index is calculated based solely on the linehaul prices of the transactions, which represents contractually defined fixed prices in the market. It does not include fuel surcharges or other accessorial charges added to the base price that reflect specific load or time-based attributes that are sometimes included in other industry indexes.

Rather than taking the index value itself, we use the percent change of the index to measure market change - typically week-over-week or month-over-month change, depending on update frequency of the index available. The percent change of the indexed value for time \( t \) is defined as \( \hat{I}_t = (I_{t-1} - I_{t-2})/I_{t-2} \). The percent change in index value, \( \hat{I}_t \), can take both positive and negative values, which allows the indexed price defined in Equation 4 to increase or decrease.
We use the index values of the preceding two periods, $I_{t-1}$ and $I_{t-2}$, because index values are often reported for the previous period. That is, in time $t$, $I_{t-1}$ would be the most recently available index value. Figure 4.1 shows the Morgan Stanley Index from 2015 to 2020 and the month-over-month percent change.

Figure 4.1: Cass Truckload Linehaul Index Freight Index, M/M percent change

The offered indexed price for loads during period $t$, $I_{nt}$, is defined by increasing or decreasing the indexed price calculated for the previous period, $t - 1$, by an amount determined by the percent change of the index in that period as well. The initial period’s indexed price is based off of the fixed price, $F_0$. Thus, we have

\[
\begin{align*}
I_{n0} &= BM_{i,j,t_0}, \quad t = t_0 \\
I_n &= I_{n-1}(1 + I_t), \quad t > t_0
\end{align*}
\]

(4)

4.3.2 Collar:

Shippers may choose to implement a collar, or an upper and lower bound on the amount the indexed price can fluctuate. One disadvantage of an indexed price is that it adds price uncertainty for shippers and carriers alike. As a safeguard against large swings in the market, a collar can dampen the price volatility: $I_t \in [\underline{c}, \overline{c}]$ and here we assume symmetry $c = |\overline{c}| = |\underline{c}|$:

\[
Offered\ Price,\ collar = \begin{cases} 
I_t, & \underline{c} \ast BM_{i,j,t_0} \leq I_t \leq \overline{c} \ast BM_{i,j,t_0} \\
\underline{c} \ast BM_{i,j,t_0}, & I_t < \underline{c} \ast BM_{i,j,t_0} \\
\overline{c} \ast BM_{i,j,t_0}, & I_t > \overline{c} \ast BM_{i,j,t_0}
\end{cases}
\]

(5)
Where, the lower bound, \( c < 0 \), and the upper bound, \( \bar{c} > 0 \), represent percentages below and above which the shipper and carrier do not want the indexed load cost change to exceed due to indexing. Note that we apply the collar symmetrically, meaning the indexed price is allowed to increase or decrease by the same amount. An alternative approach, is an escalator design; a special case of the collar design where the indexed price is used as a price increase to incentivize carriers to accept loads but the indexed price does not decrease below the initial price, \( F_0 \). That is, \( c = 0 \) and \( \bar{c} \) is some positive value or is infinite in the case where the indexed value is unbounded from above.

### 4.3.3 Initialization price

Finally, the shipper and carrier agree to an initialization price for the Indexed contract in \( t_0 \), \( F_0 \), as described in Equation 4. Shippers want to know how well each policy and design performs if the choice of initialization price, \( F_0 \) upfront is too high or too low. Determining what that price should be is difficult due to market uncertainties. Thus, one of our design choices is \( \alpha \), or how \( F_0 \) relates to the going lane benchmark price at \( t_0 \):

\[
F_0 = \alpha \cdot BM_{i,j,t_0}
\]  

(6)

Where \( \alpha > 0 \) and is a multiplier on \( BM_{i,j,t_0} \), the benchmark price on lane \((i,j)\) at time \( t_0 \).

We test the relative performance of these contract designs and policies on empirically derived demand. We use an extensive transaction dataset to model the best design choices as a shipper considers the trade-offs between expected costs and carrier acceptance rate performance.

### 5 Methodology

In this section, we describe the price mechanisms and policies we test, our model selection methodology, and the model specifications we choose.

#### 5.1 Indexed price mechanism design and contract policies

We utilize a highly detailed empirical dataset that contains 2.2 million long-haul (i.e., loads that moved a distance greater than 250 miles\(^1\)), dry van loads, all of which originate and terminate in the continental United States. The observations represent the load tenders and carrier acceptance and rejection decisions for TL loads for 47 shippers and their 308 asset and 74 non-asset primary carriers plus 430 backup carriers. The data span over four years (2015-2019). This encompasses a Soft market period from the beginning of the dataset to July 2017 and from January 2019 onward and a Tight market period from July 2017 to January 2019 (see Acocella et al. (2020) for market period justifications).

Acocella et al. (2021a) describe promising segments of a shipper’s freight and network to apply market-based contracts. We split our data according to these segments (see Appendix A), apply each indexed design, predict carriers’ acceptance behaviors for each, and determine the expected costs and primary carrier acceptance rate under each policy. Thus, we demonstrate the best indexed-based contract design for each segment.

In addition to the variables and data described above, we have our four empirically trained carrier acceptance prediction models of the probability each load is accepted for each primary carrier service type and market condition combination: (i) Asset primary carriers in a Tight market, (ii) Asset primary carriers in a Soft market, (iii) Non-Asset primary carrier in a Tight market, and (iv) Non-Asset primary carrier in a Soft market. Moreover, we model backup carrier acceptance

---

\(^1\)We use this long-haul distinction because pricing structures for the alternative, short-haul moves, differ from those we consider and discuss in this research.
probabilities as well. The method to train, validate, and test these models is described in detail in Section 5.2.1. The resulting probabilities are inputs to the expected cost formulation of Equation 2, which we calculate for each carrier service type and market condition combination.

### 5.2 Carrier acceptance decision model

We build our empirical carrier acceptance decision models on load transaction data described in Section 5.1. First, we must develop an accurate model to predict the probability that a carrier accepts or rejects a load, given the characteristics of the load, lane, shipper tendering behavior, and price. We must also consider whether the carrier is a primary or backup carrier, as the agreements, incentives, and relationships differ between shippers and their primary carriers as compared to their backup carriers (see Section 5.2). As such, we develop five distinct carrier acceptance response models: four for each of the primary carrier service type-market combination, and a fifth for backup carriers.

To do so, we split the data into one set of the load tenders from a shipper to its primary carrier on the lane (regardless of if the primary carrier accepts the load), and a second set of all load tenders from a shipper to backup carriers on the lane, which only occur if the primary carrier has rejected the load. We then model the probability that a load is either accepted or rejected (i.e., a binary outcome) by a carrier based on multiple associated load, lane, and carrier characteristics. For brevity, rather than describing each in detail here, we refer the reader to the working paper, Acocella et al. (2021a) for a detailed description of each variable we include in our model. Table 5.1 below describes each one.

#### Table 5.1: Carrier Acceptance Model Input Variables

<table>
<thead>
<tr>
<th>Variable*</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spot Rate</td>
<td>Difference between spot price and offered price, as percentage of offered price</td>
</tr>
<tr>
<td>Differential (SRD)*</td>
<td></td>
</tr>
<tr>
<td>Lane demand</td>
<td>Percent of the preceding 4 weeks in which loads are tendered to primary carrier by shipper on the lane</td>
</tr>
<tr>
<td>cadence</td>
<td></td>
</tr>
<tr>
<td>Lane demand</td>
<td>Four-week rolling average week-over-week percent difference in tendered volume from shipper to primary carrier on the lane</td>
</tr>
<tr>
<td>volatility</td>
<td></td>
</tr>
<tr>
<td>Load surge</td>
<td>Load’s rank within the week relative to the expected weekly volume as a percentage</td>
</tr>
<tr>
<td>category</td>
<td></td>
</tr>
<tr>
<td>Asset carrier</td>
<td>Number of tractors in the carrier’s fleet, categorized</td>
</tr>
<tr>
<td>fleet size*</td>
<td></td>
</tr>
<tr>
<td>Distance*</td>
<td>Number of miles between origin and destination locations</td>
</tr>
</tbody>
</table>

*Variables that are included in both the primary carrier and the backup carrier acceptance rate models. Otherwise, only included in the primary carrier model.

#### 5.2.1 Imbalanced data

Primary carriers are expected to maintain high acceptance rates. While shippers’ service level requirements differ, carriers may be expected to accept at least 90, 95, or even 99% of loads tendered to them. Of course, this level of service may drop due to capacity availability or contract prices becoming out of date as the freight market changes over the course of the contract\(^2\).

\(^2\)It is typical for shippers to expect primary carriers to flex up capacity availability and maintain this level of tender acceptance for “surge” volume up to 10% - sometimes more - above the awarded volume.
In our dataset that spans multiple market cycles, the average acceptance ratio of primary carriers is 82%. In other words, the frequency of observations in our primary carrier dataset in which the load is rejected is much lower than the frequency of the observations in which the load is accepted. Thus, our segmented primary carrier dataset is imbalanced. Predictive classification models trained on imbalanced data are often biased toward the majority class, or the class containing most observations; here, the “accepted” class. In our case, the imbalanced problem only affects the primary carrier model, as the backup carrier dataset is balanced.

The class imbalance problem has been addressed in the literature (e.g., He and Garcia (2009); Sun et al. (2009)). It is found in many real-world contexts such as credit card fraud detection (Brown and Mues, 2012; Arminger et al., 1997), corporate bankruptcy prediction (Altman, 1968), and rare disease diagnosis (King and Zeng, 2001). Models based on imbalanced data suffer from high error rates in classifying the minority class because there are many more instances of majority class than minority class observations. For example, a model trained to maximize accuracy (the percent of all observations that are correctly classified) on an imbalanced dataset with a ratio of majority to minority class of 80:20 could achieve 80% accuracy by simply predicting all observations to be in the majority class without even considering the observations’ features. Thus, common model performance measures such as accuracy are ill-suited for imbalanced datasets, particularly when good predictive performance is desired for both majority and minority classes.

5.2.2 Algorithmic approach

To obtain models with high predictive performance on our imbalanced data we compare a set of classification algorithms and choose the one and its corresponding best-tuned hyperparameter values (Ye et al., 2013; Sun et al., 2017; Lin et al., 2017) for our primary (and backup) carrier load acceptance decisions. See Appendix B for more detailed discussion of the different approaches to dealing with imbalanced data. We test a set of algorithmic methods commonly used in the literature for binary classification on imbalanced data: Logistic regression (with and without penalty terms), clustering with k-Nearest Neighbors, and ensemble methods, in particular random forests.3 For each model described below, our input variables are those described in Section 5.2 and we model Asset primary carriers and Non-Asset primary carriers in each market period separately.

**Logistic Regression:** Logistic regression models are a common choice for binary classification. One powerful advantage of the logistic regression model is it allows the modeler to isolate the relationship between each input variable and the binary response variable. However, many real-world data may not abide by the assumptions required to apply (log) linear models. We expect other models may better represent our empirical data with better predictive performance.

**Regularization:** In addition to the logistic regression model without penalty terms discussed above, we consider three regularization extensions: logistic regression with an L1, or linear penalty term (Lasso); logistic regression with an L2, or quadratic penalty term (Ridge), and logistic regression with a linear combination of L1 and L2 penalty terms (Elastic Net). These three models tend to improve the predictive performance of the non-penalized logistic regression model by minimizing or eliminating the influence of unimportant input variables (Tibshirani, 1996; Hastie et al., 2016).4

---

3We do not test another common classification approach, the support vector machine (SVM) algorithm, because of long run times associated with SVM and large datasets - i.e., dataset larger than tens of thousands of records, much smaller than our primary carrier tenders dataset.

4The L1 penalty forces the estimated coefficient of some input variables to zero in the regression error minimization calculation, resulting in a simpler model and, in many cases, better predictive performance than the non-penalized regression model. On the other hand, the L2 penalty shrinks the coefficients of unimportant input variables, but does not allow them to equal zero. This is advantageous when the model includes input variables that are highly correlated, but still offer important information about the data. The L1 penalty would choose one of the variables to keep and force the coefficients of the others to zero, whereas the L2 penalty would retain all variables but reduce...
Clustering: We consider a clustering technique commonly used in classification problems, k-Nearest Neighbors (kNN). The model makes no assumptions on relationships between variables and is generally easy to interpret (Hastie et al., 2016).

Ensembles: Finally, we consider an ensemble classification technique, a random forest, which is commonly used due to its good classification performance across many data types. In fact, random forests have been shown to handle imbalanced data well (Breiman, 2001), particularly when the imbalance rate is below 15% (Xia et al., 2016; Daho et al., 2014; Perry and Bader-El-Den, 2015; Zhu et al., 2018).

5.2.3 Scoring metric

While some classification metrics described in Appendix B.3 are better suited for the imbalanced problem than others, they still do not capture the specific problem at hand. In our case, good classification performance of each class is not the most important outcome of the prediction model. Our model is to be used as an input to an expected cost formula. As such, we need a model that has good performance predicting the probability an observation is in each class.

There has been limited literature focusing on probability estimates on imbalanced data, however there has been work studying how to obtain good probability predictions in general (Niculescu-Mizil and Caruana, 2005; Zadrozny and Elkan, 2002). For probability predictions, typically a Brier Score is used (Brier, 1950). The Brier score, $BS$ is the mean square error between the predicted probability an observation is in the positive class, $p_n^+$, and the actual outcome, $o_n^+ \in \{0,1\}$. A 1 indicates the observation is in the positive class and a 0 indicates it is not: $BS = (1/N) \sum_{n=1}^{N} (p_n^+ - o_n^+)^2$, where $N$ is the total number of observations in the dataset.

The Brier Score takes values between 0 and 1. Better models have lower Brier Scores. However, this standard Brier Score does not apply well to imbalanced datasets as it does not capture whether a model is biased toward the majority class; The formulation averages the error term over all observations, $N$. A model that predicts the probability of observations in the majority class well (i.e., low probability predictive error of majority class observations) will still have a low Brier Score even if it poorly predicts the minority class, because there are fewer observations in this class to outweigh the low error on the majority class. To overcome this issue, Wallace and Dahabreh (2014) propose a stratified Brier Score, where we calculate class-specific Brier Scores:

$$BS^+ = \frac{1}{N^+} \sum_{n=1}^{N^+} (p_n^+ - o_n^+)^2$$

$$BS^- = \frac{1}{N^-} \sum_{n=1}^{N^-} (p_n^- - o_n^-)^2$$

The authors show that the stratified Brier Scores, $BS^+$ and $BS^-$, are effective measures of how well a range of different supervised classifiers predict class probabilities for both majority and minority classes of known imbalanced datasets. Thus, we adopt the stratified Brier Scores to evaluate and compare carrier load acceptance prediction models: logistic regression, regularized logistic, kNN, and random forest. We describe our algorithmic approach including model hypertuning and model selection in the following section.

---

the relative importance of those that contribute to higher error. The Elastic Net regularization allows the benefits of each L1 and L2 and balances the two penalty terms with a regularization parameter, $\lambda$. 

13
5.3 Model hypertuning and comparison

Because no single model performs best on all datasets - imbalanced or otherwise - we adopt a cross validation method to determine the best model to predict the probability a load is accepted by a primary carrier. That is, the model which results in the lowest combination of stratified Brier Scores. Details of our hypertuning process are in Appendix C. Results of the model selection for the Asset primary carriers in Tight market conditions are reported in Table 5.2.

Table 5.2: Hypertuned prediction model Brier Scores: Asset, Tight market

<table>
<thead>
<tr>
<th>Model</th>
<th>Standard BS</th>
<th>Majority BS⁺</th>
<th>Minority BS⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.178</td>
<td>0.178</td>
<td>0.175</td>
</tr>
<tr>
<td>L1 Regularization*</td>
<td>0.178</td>
<td>0.177</td>
<td>0.181</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td>0.078</td>
<td>0.032</td>
<td>0.282</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.047</td>
<td>0.036</td>
<td>0.097</td>
</tr>
</tbody>
</table>

*L2 and Elastic Net results excluded because L1 regularization always outperforms L2 in the hypertuning step, and Elastic Net with regularization parameter, $\lambda = 1$ (corresponding to L1), is best.

The results show that the hypertuned random forest outperforms all other hypertuned models in predicting the probability a load is accepted by a primary carrier due to its low majority and minority class Brier Scores. We apply the same method for the dataset of Asset carriers in Tight markets reported above to all other carrier-market condition combinations. The relative performance of the random forest holds in all cases. Thus, we use the balanced random forest as the empirically derived probabilistic model for our expected cost equations.

Although the backup carrier data are balanced, we use a similar backup carrier acceptance model selection method comparing the models’ to ensure good predictive performance. The backup carrier input variables include the load’s spot rate differential, origin and destination region indicator variables, and distance. Recall that other shipper tendering behavior, lane, and load characteristics included in the primary carrier model do not pertain to non-contracted, backup carriers. The resulting best model for both carrier segments is the random forest with uniform class weighing.

5.4 Design and policy choices tests

Next, we test our contract design and policy choices. For each design choice parameter (i.e., choice of $\alpha$ to determine initial fixed price and $c$ to determine size of the collar) we apply a range of values for each parameter for each network segment, market condition, and carrier service type. We calculate the resulting primary carrier acceptance (PAR) (Eq. 3) and expected cost per mile (Eq. 2) for all parameter values.

Recall, $\alpha$ determines how much above or below the lane benchmark price at which the indexed pricing initiates and is a multiplier on that lane benchmark price. Therefore, a value of 1 corresponds to using the actual lane benchmark price to initialize the indexed contract. We range $\alpha$ from 0.01 to 2.5. For the collar parameter, we range $c$ from 1% to 100%. This allows the indexed price to fluctuate up or down around the fixed benchmark by that capped amount.

The “best” value of $\alpha$ and $c$ depend on a shipper’s cost and performance tolerances. Therefore, we report the PAR vs. expected cost per mile curves. These curves show the rate at which PAR increases with changes in expected costs (resulting from dynamic indexed contract prices and reliance on backup or spot options) as we change the design choice parameters - in other words, the marginal increase in PAR for an increase in expected costs. In this way, we capture diminishing
returns and assist the shipper in determining the design that captures the most benefit for a cost it deems reasonable.

Once we have the best design choice parameter values, we compare the three policy choices to one another. That is, we compare the PAR and expected costs if the shipper implemented the Indexed policy ($\pi_1$) or the Surge only policy ($\pi_2$) as compared to the traditional Fixed-price policy ($\pi_0$).

6 Results

A shipper will choose an Indexed contract that offers lower expected cost or higher primary carrier acceptance, or both, as compared to the status-quo Fixed price contract. In this section we discuss the contract design parameter conditions for each segment and carrier-market condition combination in which the Indexed contracts results in a Pareto improvement over the Fixed price contract policy. In other words, the segments and their corresponding Index-based contract policy in which both parties are at least as well off and at least one is better off. This occurs when either the shipper experiences lower expected cost or increased PAR and the carrier sees increased expected revenue. In this section, we summarize the key findings.

We present the particular Indexed contract designs for two scenarios:

Case I: The carrier’s expected revenue is greater than or equal to that of the benchmark Fixed-price contract and the shipper’s expected cost is less than or equal to the benchmark Fixed-price contract and the shipper sees higher or the same primary carrier acceptance rate (PAR) relative to that of the benchmark Fixed-price contract;

Case II: The carrier’s expected revenue is greater than or equal to that of the benchmark Fixed-price contract and the shipper sees higher or the same primary carrier acceptance rate (PAR) relative to that of the benchmark Fixed-price contract at some reasonable expected cost increase relative to the benchmark Fixed-price contract. A reasonable increase in expected cost increase is set to 10%, which reflects what a shipper typically pays for its backup carriers between both Soft and Tight markets (see Aemireddy and Yuan (2019)).

Finally, the choice of implementing a collar and corresponding $c$ value impacts the primary carrier acceptance, shipper’s expected cost, and carrier’s expected revenue to a small extent. At values of $\alpha < 0.5$, the collar does not impact the results to a large extent. At values of $\alpha > 1.1$, the collar threshold value is hit more often. Thus, at the optimal values of $\alpha$ from 0.9-1.1, presented in Tables 6.1 and 6.2, the resulting acceptance rates, costs, and revenues are not impacted by different values of $\alpha$. As such, we do not report values of $c$ in these summary tables.

The summary of benefits for Case I is reported in Table 6.1 all with no collar on the price. Indexed-based contracts will be agreed to under the Case I conditions only in Tight markets on three of our segments: High Volatility and Moderate and Low Cadence Lanes. In all cases, the initialization price should be set to about 5-10% below the lane benchmark price.

| Table 6.1: Indexed contract design improvements, Case I |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Segment                        | Market Condition| Carrier Type    | Exp. cost reduction | PAR increase | Exp. rev. increase | $\alpha$  |
| High Volatility                | Tight           | Asset           | 1%                | 3%            | 8%              | 0.95           |
|                                |                 | Non-Asset       | 5%                | 2%            | 1%              | 0.90           |
| Mod. Cadence                   | Tight           | Asset           | 1%                | 3%            | 8%              | 0.95           |
|                                |                 | Non-Asset       | 5%                | 4%            | 3%              | 0.90           |
| Low Cadence                    | Tight           | Asset           | 1%                | 4%            | 9%              | 0.95           |
On High Volatility lanes, the shipper sees the best reduction in expected cost (5%) with a 2% increase in PAR by implementing the Indexed contract with Asset carriers - who see a 1% increase in expected revenue - rather than Non-Asset carriers. On Moderate Cadence lanes, the Non-Asset carrier offers better expected cost reduction (5%) and PAR increase (4%) than the Asset carrier. Finally, on Low Cadence lanes, only Indexed contracting with Asset carriers results in the conditions for Case I.

Next, we summarize the outcomes for Case II in Table 6.2. Case II differs from Case I in that here, the shipper accepts an increase in expected cost up to the average backup premium of 10%, for an increase in PAR. We expect that in doing so, the shipper will have a higher increase in PAR than in Case I. Similar to Case I, here in order for the carrier to agree to the contract terms, its expected revenue should be at least as good as that of the benchmark Fixed price contract.

Table 6.2: Indexed contract design improvements, Case I

<table>
<thead>
<tr>
<th>Segment</th>
<th>Market Condition</th>
<th>Carrier Type</th>
<th>Exp. cost increase</th>
<th>PAR increase</th>
<th>Exp. rev. increase</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Volatility</td>
<td>Tight</td>
<td>Asset</td>
<td>1-4%</td>
<td>8-11%</td>
<td>19-28%</td>
<td>1-1.05</td>
</tr>
<tr>
<td>Mod. Cadence</td>
<td>Tight</td>
<td>Asset</td>
<td>4-7%</td>
<td>6-9%</td>
<td>17-26%</td>
<td>1-1.05</td>
</tr>
<tr>
<td>Low Cadence</td>
<td>Tight</td>
<td>Asset</td>
<td>3-9%</td>
<td>6-10%</td>
<td>17-33%</td>
<td>1-1.10</td>
</tr>
</tbody>
</table>

The results suggest that Case I and II only occur during Tight market conditions. This is because the freight market index - and resulting index-based price - decreases during Soft markets. In order for the carrier’s expected revenue to increase relative to the Fixed benchmark priced contract in the Soft market conditions, the initialization price would have to be so high relative to the benchmark price that the shipper’s expected cost increases well above our 10% threshold. In this case, our models offer carriers a starting point for discussions with shippers to consider their willingness to accept higher expected costs for higher load acceptance from contracted carriers with which a relationship is in place and a history of demonstrated performance.

To put these numbers into perspective, say a shipper’s total annual spend is around $100 million. This is in par with medium sized shippers in the US. Even a 1% reduction in costs would result in $1 million dollars in transportation savings. As one shipper put it, “$1 million dollars saved on transportation is $1 million we can spend on our actual product.”

7 Case Study: Index-based TL contracts, pilot experiment in practice

We work with a large US agricultural shipper to implement a pilot of index-based contracts in its long-haul TL network. For purposes of anonymity, we refer to the shipper as AgCorp. We use their load-level transaction and carrier acceptance/rejection data for two purposes: to (i) validate our carrier behavioral models, and (ii) quantify the causal effect of index-based pricing in the freight transportation context.

7.1 Pilot implementation and data description

During the height of the Covid-19 pandemic in the US - summer and fall of 2020 - AgCorp ran its annual TL transportation procurement event. Due to concerns of elevated contract price bids it received and carrier acceptance levels plummeting - similar to the rest of the shipper community during this time - AgCorp implemented index-based contracts with two of its Non-Asset carriers.

The contract policy and design choices correspond to those we model and test in the preceding sections. The contracts apply indexed pricing to all of the loads (i.e., policy \( \pi_1 \)) on the ten chosen pilot lanes, all originating from a major inbound city in its network. They use the month-to-month
actual dollar change in DAT’s national TL index to update the indexed contract price monthly with no collar. They choose the initializing price, $F_0$ to be the lane-level benchmark price defined by DAT at the start of the contract, $BM_{i,j,t_0}$ (i.e., $\alpha = 1$). Finally, the contracts are implemented during a Tight market period.

The pilot lane characteristics are characterized by the segments discussed in the preceding sections: award volume, demand cadence and volatility, and surge patterns. On average, the award volume is about 1 load per week. Average volatility is 1.47 (i.e., on average, the week-to-week tendered volume is 147% more than the awarded volume of 1 load per week. For example, 2-3 loads per week). Cadence and Surge volume category splits are reported in Table 7.1. The average length of haul is 1,224 miles.

Table 7.1: Summary of Pilot Lanes

<table>
<thead>
<tr>
<th>Awd. Volume (# loads/wk)</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatility</td>
<td>1.47</td>
</tr>
<tr>
<td>0%</td>
<td>15.7</td>
</tr>
<tr>
<td>Cadence Category (% of lanes)</td>
<td>13.9</td>
</tr>
<tr>
<td>25%</td>
<td>14.8</td>
</tr>
<tr>
<td>50%</td>
<td>20.4</td>
</tr>
<tr>
<td>75%</td>
<td>35.2</td>
</tr>
<tr>
<td>100%</td>
<td>49.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surge Category (% of lanes)</th>
<th>4.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% above award</td>
<td>0</td>
</tr>
<tr>
<td>10-20% above award</td>
<td>46.3</td>
</tr>
<tr>
<td>&gt;20% above award</td>
<td></td>
</tr>
</tbody>
</table>

| Avg Lane Distance (miles)   | 1,224 |

AgCorp provides detailed transaction-level data of every load it tenders to each of its long-haul, TL carriers from October 2018 through April 2020. The data consist of each carrier’s accept or reject decision and subsequent tendering if the load is rejected, the load price, and lane details. Moreover, we are provided with the award information: expected volume and price of each awarded carrier for the three bid events AgCorp conducted (2018, 2019, and 2020). As the pilot indexed contract program (“treatment”) is implemented in October 2020, we have data that represents pre-and post-implementation shipper and carrier behaviors and decisions. We also have benchmark prices for every lane from the Freight Market Intelligence Consortium (FMIC), a well-established freight analytics group.

7.2 Model validation

First, we use these pilot data to validate our carrier acceptance models - the engine of the contract design questions. To do so, we apply our trained non-asset, tight market carrier acceptance model (recall, a “weighted” random forest model using input factors listed in Table 5.1) to the loads that are tendered under the index priced contracts. This subset of the pilot study data is similarly imbalanced to the training dataset, so it is reasonable to expect our models to perform well.

In the imbalanced dataset context, a commonly used measure of the model’s ability to predict both majority and minority classes is the precision versus recall curve (Saito and Rehmsmeier, 2015; Japkowski, 2001; Kotsiantis et al., 2006), its Area Under the Curve (AUC), and their harmonic mean, the F1 score (Bradley, 1997). Precision measures the fraction of True Positive to Positive Predictions and Recall measures the fraction of the positive observations that are successfully classified (see Appendix B.3. Finally, the F1 score is the weighted average of Precision and Recall.
It takes both false positives and false negatives into account:

\[ F1 = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \]

Applying our trained model of Non-Asset carrier in Tight market acceptance to the pilot data, we get an AUC of 98.6% and an F1 score of 96.3%. From these results, we conclude that our trained model is a very good predictor of carrier acceptance for new, unseen data under an actual indexed pricing environment. This validates the results from our contract design models.

7.3 Causal effect of index-based freight contracts

Next, we utilize the experimental pilot data and a causal inference approach to isolate the effect of index-based contracts on primary carriers’ acceptance rates and costs of loads that the shipper pays. Due to the non-random assignment of carriers and lanes to indexed contracts (i.e., “treatment”) we use propensity score matching to estimate the average treatment effect and bootstrap sampling to estimate the standard errors of this two-stage optimization problem (Austin and Stuart, 2017).

7.3.1 Research question

As the indexed contracts are implemented during a Tight market period, all primary carriers’ acceptance rates decline during this time period. However, as demonstrated in Figure 7.1, the treated carriers (i.e., Pilot Carriers on Pilot Lanes) do not show nearly as dramatic a reduction, even compared to themselves on different lanes under Fixed-price contracts (Pilot Carriers on Non-Pilot Lanes). This leads us to the key research questions:

1. Is the higher maintained PAR of Pilot Carriers on Pilot Lanes in the Tight market because of the indexed contracts or due to some other confounding carrier or lane factor(s)? - e.g., lane demand patterns such as volume, volatility, cadence, or carrier service type.
2. Do Index-based contracts result in lower costs for the shipper?

Figure 7.1: Primary Carrier Acceptance Rate by pilot carrier-lane combination
7.3.2 Causal Inference Methodology

In control trials as this, it is typical to consider the impact of a policy intervention or treatment on a sample of the population where we have some individuals assigned to a test group that receives the treatment, and a control group of individuals that do not receive treatment. We could measure the average difference in the outcome of interest as the difference between these two populations. But, even if the assignment of individuals to groups were completely randomized we would potentially have uncontrollable, unmeasured, or confounding variables that introduce bias in the estimate of the treatment effect.

Instead, we can use matching techniques (Stuart, 2010). These methods pair treated observations with untreated observations that are similar across a set of potentially confounding factors. In this way, we construct a pseudo-counterfactual for each treated observation. The difference in outcome between the treated and untreated matched observations is the treatment effect. Averaging this value across all matched observations results in the average treatment effect. Figure 7.2 describes the matrix of carrier-lane-treatment conditions and how they relate as potential matches to the treated observations.

Table 7.2: Treated and Untreated Observations and Matching Potential

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>Carrier Lane</th>
<th>Pilot</th>
<th>Non-Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Pilot</td>
<td>Pilot</td>
<td>Good potential match</td>
<td>Potential match</td>
</tr>
<tr>
<td></td>
<td>Non-Pilot</td>
<td>Potential match</td>
<td>Control</td>
</tr>
<tr>
<td>Post-Pilot</td>
<td>Pilot</td>
<td>Treated</td>
<td>Good potential match</td>
</tr>
<tr>
<td></td>
<td>Non-Pilot</td>
<td>Good potential match</td>
<td>Control</td>
</tr>
</tbody>
</table>

In our pilot setup, the assignment of treatment is not randomized. AgCorp chooses specific carriers with which to implement indexed contracts, and a subset of lanes originating from an important origin location. In these contexts, Propensity Score Matching (PSM) has been demonstrated to result in unbiased, accurate treatment effects (Austin and Stuart, 2017; Stuart, 2010). Moreover, PSM outperforms pure matching methods when there are dimensionality concerns.

We choose a set of potentially confounding factors (lane and carrier characteristics) that may be influencing the observed outcome (i.e., primary carrier acceptance rate and load prices). For example, Acocella et al. (2021a) demonstrates that all else equal, Asset carriers and Non-Asset carriers have different acceptance rates and pricing concerns. The Pilot Carriers are all Non-Asset carriers. Thus, we hypothesize that carrier service type is a factor that may be influencing Pilot Carriers’ acceptance behaviors. In addition, since the Pilot Lanes all originate from a major depot area within AgCorp’s network, they may be particularly attractive lanes than other lanes, which may also be influencing Pilot Carriers’ decisions (Acocella et al., 2020). Thus, the factors we include are those described in Table 5.1 and an indicator of each origin and each destination.

The PSM algorithm is as follows. First, we model the likelihood each observation is in the treated group based on this set of potentially confounding factors. This is typically done with logistic regression. The “treatment” outcome (i.e., load is tendered under Indexed contract) of each observation is regressed on the potentially confounding attributes. The resulting likelihood is the observation’s propensity score. This propensity score effectively condenses the confounding factors into one value. Second, each treated observation is matched with an untreated observation.

5In our study, we use the Generalized Estimating Equations (GEE) method with a binomial family to account for the repeated measures of individual carriers’ decisions observed throughout the dataset.
based on similar values of propensity score; this results in a pseudo-counterfactual to compare against. Matching is done with kNN, where k=1 and observations’ nearness is defined by their propensity scores.

One further challenge with matching methods is identifying proper standard errors for the average effect, as discussed in Rubin and Thomas (1996); Hill et al. (1999). Uncertainty arises at both steps of this two-stage approach. That is, there are errors introduced with the classification method used to calculate propensity scores and with the matching algorithm. Few methods have been demonstrated to mitigate this issue and measure confidence intervals (Stuart, 2010; Terza, 2016). Sampling techniques such as bootstrapping as we employ here have had some support from the literature (Sekhon, 2008; Rose, 2018). Moreover, Lechner (2002) and Hill and Reiter (2002) show that bootstrapping outperforms other methods.

We implement a sampling technique with over- and under-sampling to account for the small sample number of pilot observations. After calculating the propensity score of all observations, we sample 200 of the treated observations with replacement and 200 with replacement of the untreated observations, as discussed in Dehejia and Wahba (2002). We match the two sets with kNN and measure the average treatment effect. We repeat this process 100 times, which results in a distribution of 100 average treatment effects. We take the average over these averages and the 95% confidence interval as our resulting causal effects.

Our research questions for this pilot study consider how well indexed contracts influence primary carriers to maintain their acceptance rates during the tight market and the costs the shipper incurs. As such, the outcomes we measure are the percent change in (1) PAR and (2) accepted price of the load between pre- and post-implementation; in other words, (1) how well primary carriers sustain lane-level acceptance rates during the tight market of 2020 and (2) the spot rate differential (SRD) between pre- and post-implementation.

Our observations are defined by all the loads tendered for a carrier-lane-week number combination. “Treated” observations are loads tendered to Pilot Carriers on Pilot Lanes in the week numbers corresponding to Oct. 2020 to Apr. 2020 (the range of our dataset corresponding to the pilot period). Untreated observations are any Carrier-Lane combination to which the loads are tendered before the pilot indexed contracts are implemented and any Carrier-Lane combinations after implementation that are not tendered to Pilot Carriers on Pilot Lanes. The outcome of interest for each observation is measured as the pre- to post-treatment percent change in (1) PAR and (2) SRD corresponding to that carrier-lane-week.

The coefficient results of the logistic regression model, which calculates each observation’s propensity score, can be used to understand how important each factor is to matching treated and untreated observations - that is, what factors contribute most to measuring similarities between treated and untreated observations. We find that the most important contributor is whether the carrier is a Pilot Carrier or not. Note that we do not include a Pilot Lane indicator in the propensity score model. This is because we all lanes into their origins, destinations, and distances. By doing so, we provide more potential matches for each treated observation while avoiding multicollinearity in the propensity score regression model.

7.3.3 Causal Inference Results

Controlling for the potential confounding factors and sampling as described in Section 7.3.2, we find that there is a causal effect of indexed contracts on primary carriers’ acceptance ratio. Specifically, the average treatment effect is 12.6 with 95% Confidence Interval [12.1-13.1]. That is, if indexed pricing were implemented for all carrier-lanes, the percent change in PAR between fixed price contracts in the soft market and indexed contracts in the following tight market period would have been 12.6 percentage points higher than it actually was.
For example, Figure 7.2 depicts the actual PAR degradation AgCorp saw in the tight market of 2020, and what it would have been if it had implemented index-priced contracts. Instead of roughly 90% acceptance in the soft markets of 2018 and 2019 dropping to 65.7% in the tight market (a 27.3% reduction), AgCorp would have seen a decrease to only 77.4% (a 12.9% reduction).

Figure 7.2: Effect of Indexed pricing on PAR

However, we do not find a statistically significant reduction in the accepted cost (SRD) of the loads as a result of the Indexed contracts. These Pilot Lanes are high volatility, low volume, and predominantly moderate cadence lanes with Non-Asset Pilot carriers in Tight markets. Recall from the results in Section 6, in order for the shipper to see lower expected costs, the Indexed contract price must be initialized below the lane benchmark price (i.e., $\alpha < 1$). In this case study, the shipper set the initialization price at each lane’s benchmark price. This suggests that the design of the contract is a major contributor to how beneficial it is for the shipper. Moreover, it underscores the importance of our study’s contributions to practitioners to identify the best contract designs and policies.

8 Research Implications and Contributions

In this research, we study how a shipper and TL carrier should design and consider implementing index-based contracts into their portfolio of transportation agreements. We demonstrate that during Tight market periods, shippers and carriers can agree to Index-based contracts for High Volatility lanes and Moderate-Low Cadence lanes in Tight freight markets that result in up to 5% lower expected costs for the shipper than the status quo Fixed price contract. This corresponds to an increase in carrier contract compliance (i.e., primary carrier load acceptance rate) of up to 4% while the carriers’ expected revenue increases by up to 9%.

To put these numbers into perspective, say a shipper’s total annual spend is around $100 million. This is in par with medium sized shippers in the US. Even a 1% reduction in costs would result in $1 million dollars in transportation savings. As one shipper put it, “$1 million dollars saved on transportation is $1 million we can put toward the actual products we sell.”

Moreover, shippers may wish to see higher increase in primary carrier acceptance than suggested above and may be willing to increase their expected costs to do so. We show that shippers can increase primary carrier acceptance up to 10% while increasing expected costs to 7-9%. This results in carrier’s expected revenue increase of 26-33%. Therefore, we conclude that Index-based truckload
freight contracts show promise for both shippers and carriers.

Finally, we isolate and quantify the causal impact of Index-based contracts in Tight market conditions on primary carrier acceptance rates and costs to the shipper in practice. The results underscore the importance of considering the design choices of the contract.

8.1 Academic and practical implications

The contributions of this study are applicable for both academic and practicing audiences. First, the supplier-buyer contracting literature makes no consideration of the types of contracts we consider here - where supply, demand, and market uncertainty require flexibility in contract compliance for both sides. More specifically, contracts that are non-binding in demand (e.g., loads tendered by a shipper) and in supply (e.g., carrier acceptance of tendered loads). We add to the literature by explicitly considering these contracts, proposing an alternative design, and presenting the ways in which both parties can benefit from these contracts. This serves as a foundation for further experimental and theoretical exploration of these dynamics.

Second, much of the freight transportation contracting literature overlooks the nuances introduced by market dynamics. Simplifying assumptions around the dynamic relationship between spot and contract prices and the nature of these non-binding contracts in practice limit the applicability of the results of previous studies. Moreover, due to difficulty in obtaining proprietary company data, the extant set of literature rarely - if ever - models behaviors and prices on empirical data. Their results often abstract away from real-world challenges. Not only do we build our models and test them on an extensive empirical dataset, but we work with a large US firm to implement index-based contracts to both validate our models and demonstrate the causal effect of them in a real-world setting. Thus, we contribute empirical research to the contract design literature.

The practical contributions of this work stem from the growing interest in index-based contracts. Both buyers and suppliers of truckload transportation seek contract designs to balance the performance and cost trade-offs discussed here. However, the design, strategy, and segmentation choices considered in this research have not been explored to this extent by practitioners. Moreover, our results suggest that certain design decisions - specifically the initialization price - impact the actual effectiveness of the contract. We identify and formulate a set of contract parameters to consider and provide recommendations on how practitioners should approach the design and implementations challenges. We develop models that individual shippers can apply to their businesses to make the performance and cost trade-offs that apply best to their own transportation goals.

The contributions can be generalized to other contexts in which flexible contracts lead to an interest in dynamic pricing. In these settings, we offer a segmentation strategy that applies to demand patterns for any sector, and recommendation for how to construct and design the contracts.

An interesting stream of future research could take an analytical, game theoretic approach to our context of market-based pricing under supply, demand, and market uncertainties. However, the impact of these compounding uncertainties would be difficult to capture and likely lead to intractable solutions. Our empirical approach offers suggestions of the most important aspects to retain in future formulations or related problems.

References


Acocella, A., Caplice, C., and Sheffi, Y. “Elephants or Goldfish?: An empirical analysis of


Hastie, T., Tibshirani, R., and Friedman, J. Data mining, inference, and prediction. In *The Elements of Statistical Learning*, 2016. 2nd ed.


### Appendix A  Segmentation

#### Figure A.1: Lane Segments: Volume

<table>
<thead>
<tr>
<th>Segment</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>3177</td>
<td>72</td>
<td>76</td>
</tr>
<tr>
<td>Mean</td>
<td>26.17</td>
<td>385</td>
<td>884.99</td>
</tr>
<tr>
<td>StDev</td>
<td>50.19</td>
<td>56.48</td>
<td>430.88</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>301</td>
<td>502</td>
</tr>
<tr>
<td>50%</td>
<td>6</td>
<td>372.5</td>
<td>709</td>
</tr>
<tr>
<td>Max</td>
<td>299</td>
<td>499</td>
<td>2706</td>
</tr>
</tbody>
</table>

![Histogram of Lane Volume](image1.png)

#### Figure A.2: Lane Segments: Volatility

<table>
<thead>
<tr>
<th>Segment</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>37</td>
<td>1414</td>
<td>1828</td>
</tr>
<tr>
<td>Mean</td>
<td>0.108</td>
<td>0.633</td>
<td>1.81</td>
</tr>
<tr>
<td>StDev</td>
<td>0.066</td>
<td>0.214</td>
<td>0.635</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
<td>0.20</td>
<td>1</td>
</tr>
<tr>
<td>50%</td>
<td>0.121</td>
<td>0.644</td>
<td>1.70</td>
</tr>
<tr>
<td>Max</td>
<td>0.193</td>
<td>0.999</td>
<td>6.20</td>
</tr>
</tbody>
</table>

![Histogram of Lane Volatility](image2.png)

#### Figure A.3: Lane Segments: Cadence

<table>
<thead>
<tr>
<th>Segment</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>1024</td>
<td>492</td>
<td>1809</td>
</tr>
<tr>
<td>Mean</td>
<td>0.357</td>
<td>0.634</td>
<td>0.904</td>
</tr>
<tr>
<td>StDev</td>
<td>0.107</td>
<td>0.066</td>
<td>0.092</td>
</tr>
<tr>
<td>Min</td>
<td>0.25</td>
<td>0.503</td>
<td>0.75</td>
</tr>
<tr>
<td>50%</td>
<td>0.33</td>
<td>0.643</td>
<td>0.927</td>
</tr>
<tr>
<td>Max</td>
<td>0.5</td>
<td>0.748</td>
<td>1</td>
</tr>
</tbody>
</table>

![Histogram of Lane Cadence](image3.png)
Appendix B  Approaches for Modeling with Imbalanced Data

B.1  Pre-processing

The first common pre-processing technique is over-sampling. Either minority class samples are repeated and added to the training dataset or synthetic observations are created through interpolation of neighboring observations in the minority class (Chawla et al., 2003; He et al., 2008). The other common pre-processing approach is under-sampling, where observations of the majority class are removed until class proportions are balanced (Khoshgoftaar et al., 2007). Researchers have shown that a combination of over-sampling and under-sampling can be an effective method to deal with imbalanced data.

However, there are a number of issues that may result from such sampling techniques. First, the optimal class ratio may not be 1:1; Weiss and Provost (2003) finds that the optimal ratio depends on the data characteristics. Further, over-sampling may unintentionally amplify noise contained by the minority class, and if a synthetic data simulation technique is chosen, the best method to do so may not be trivial. Over-sampling has also been found to overfit training data (Chawla et al., 2003). On the other hand, under-sampling may lead to unintentionally removing majority class data points containing important information. However, Drummond and Holte (2003) find that the drawbacks of under-sampling outweigh those of oversampling.

B.2  Cost-sensitive Learning

Each classification model attempts to either maximize a likelihood estimator (MLE) or minimize a loss function to determine each observation’s class. It has been shown that adjusting the class weight in these MLE or loss function equations can improve minority class predictions for imbalanced data. It is common to assign weights by inverse class frequency (Huang et al., 2016; Wang et al., 2017) or by a smoothed, inverse square root of class frequency (Mikolov et al., 2013; Mahajan et al., 2018; Cui et al., 2019). In some model specifications, class weighting reduces to - or can even outperform - resampling techniques (King and Zeng, 2001; Xie and Manski, 1989; Scott and Wild, 1986; Amemiya and Vuong, 1987).

B.3  Scoring metric

Finally, some metrics commonly used to determine model performance are inadequate for imbalanced data classification problems.

Accuracy, for example, does not capture how well a biased model predicts the minority class (here, rejected loads). Measures such as recall and precision are similarly poor measures of the minority class. Recall is the fraction of all truly positive observations that are correctly classified: \( \text{Recall} = \frac{TP}{TP + FN} \), where \( TP \) is the true positive predictions, or the number of observations predicted to be in the positive class that are actually in the positive class, and \( FN \) is the false negative predictions, or the number of observations predicted to be in the negative class but are actually in the positive class. Precision is the fraction of positive predicted observations that are actually positive \( \text{Precision} = \frac{TP}{TP + FP} \), are good for measuring how well the model predicts the positive (majority, or “accepted”) class.

F-scores are a popular choices for imbalanced data, as they are a set of metrics that create a linear combination of precision and recall (Japkowsk, 2001; Kotsiantis et al., 2006). Another widely used method that is promising for imbalanced data problems is the receiver operating curve (ROC), which plots \( FP \) on the x-axis and \( TP \) on the y-axis and its corresponding area under the curve (AUC). These tools allow the user to choose the model parameter values that make the best trade off between misclassification types and do not place more emphasis on one class over the other (Bradley, 1997).
Related are cost-sensitive scorers, where rather than including the cost of misclassification in the error term during the model training, those costs are applied to the model’s ability to classify each class separately and, weight them based on the known costs of misclassification to achieve a single model score. However, this requires knowledge of costs misclassifying each class.
Appendix C  Model Hypertuning Process

We split each of our primary carrier datasets (defined by one of four carrier service type {Asset, Non-asset} - market condition {Tight, Soft} combinations into a training and validation set (70% of the primary load tenders) and a test set (30%), stratified so that the proportion of majority ("accepted") to minority ("rejected") class samples in the original, full dataset is maintained in each of the resulting sets. Each model under consideration is defined by a set of hyperparameters. We consider only the set of hyperparameters for each model relating to issues that arise from the imbalanced data problem.

All models take a class weight parameter, which describes how the model weighs prediction errors of each class in its corresponding cost function minimization. The class weight parameter takes value of either “uniform”, where there is no re-weighting of classes, or “balanced”, where class weights are defined as the inverse of class frequencies so that the minority class is given higher weighting to increase its contribution to total error (King and Zeng, 2001) and Xie and Manski (1989).

The logistic regression model and the first two regularizations (L1 and L2) do not have additional relevant hyperparameters to tune. The elastic net model includes a regularization parameter, $\lambda \in [0, 1]$, which defines the linear combination of the L1 and L2 penalty terms. We test discrete values of $\lambda \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$, where a value of 0 corresponds to the L2 penalty regularization, and a value of 1 corresponds to the L1 penalty regularization.

The kNN model takes parameters $k$, the number of neighboring observations to consider in classification prediction, and observation weights, measuring the weight each observation has on the classification decision based on its distance from that observation. We test the set of k values: $k \in \{5, 10, 25, 30, 35, 40, 45, 50, 60, 75, 100, 125, 200\}$. The observation weights can be either uniform, where all points in each neighborhood are weighted equally, or weighted, where all points are weighted by the inverse of their distance from the observation point.

Finally, the random forest takes parameters for the number of estimators (i.e., single trees) to construct. We test the performance of 10, 50, 100, and 200 estimators. At each node split we allow the algorithm to consider all input features.

For each combination of parameter values and models, we calculate the 5-fold (stratified) cross-validated majority Brier Score, $BS^+$, minority Brier Score, $BS^-$, and the standard Brier Score. We choose the set of parameter values that results in the lowest combination of majority and minority Brier Scores. This best parameter combination for each model constitutes the best-tuned, or hypertuned model.

We find that all models are best tuned by a balanced class weight. The best regularization parameter for the elastic net model takes value of $\lambda = 1$ (i.e., the special case corresponding to the L1 regression model). The best trade-off between Brier Scores for the kNN model results in $k = 75$ and uniform observation weights. The best-tuned random forest uses 100 trees. We then apply each best-tuned model to the test data and again calculate the majority and minority Brier Score. We choose the single hypertuned model that results in the best combination of majority and minority Brier Scores on the test data.

While the random forest model does not offer an analytical relationship between input variables and the predicted outcome, we can measure the relative importance of each input feature based on how many important node splits are made in each sub tree for each variable. This is measured by the decrease in Gini impurity score that results from splits on that variable. The most important input feature to correctly predict if a load is accepted is its offered price spot rate differential, followed by lane cadence, origin, destination, carrier service type and size, and lane volatility. These results also correspond to the econometric model results in Acocella et al. (2021a).