Characterization and Short Term Forecasting of the US Long Haul Truckload Spot Market

by

Shraddha Rana

B. Tech. with Honors, Civil Engineering
Indian Institute of Technology, Bombay, 2017

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN TRANSPORTATION

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2019

Signature redacted

Signature of Author: ________________________

Department of Civil and Environmental Engineering
May 17, 2019

Signature redacted

Certified by: ________________________

Chris Caplice
Senior Research Scientist, MIT Center for Transportation and Logistics
Thesis Supervisor

Signature redacted

Accepted by: ________________________

Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Characterization and Short Term Forecasting of the US Long Haul Truckload Spot Market

by

Shraddha Rana

Submitted to the Department of Civil and Environmental Engineering on May 17, 2019
in Partial Fulfillment of the Requirements for the Degree of Master of Science in Transportation

Abstract

Accurate forecasting of transportation costs is a key step in logistical planning. It helps buyers and sellers of transportation services make better decisions at all stages of a supply chain, thus creating a significant need to develop forecasting techniques that give useful results. First, we study the truckload market in the US by defining indicators that capture the market characteristics. Then we explore techniques for making short term weekly forecasts for truckload spot market rates at a national level and of selected 3-Zip origin regions in the USA. Short term spot rate forecasts help with making operational decisions, estimating budget for shippers, and cash flow for carriers. But making frequent forecasts for volatile time series such as truckload spot rates comes with its challenges. We solve the problem using four models: Naïve, Moving Average, Auto Regressive Integrated Moving Average, and Feed-Forward Neural Networks. Additionally, we employ concept drift handling techniques to re-train the models regularly with new information to account for changes that may appear in the underlying data structure over time. Finally, we draw inferences from the MAPEs of the models and comment on their merit.

Thesis Supervisor: Chris Caplice
Title: Senior Research Scientist

Acknowledgements

I would like to extend my gratitude towards my thesis advisor, Dr. Chris Caplice, for his continuous motivation and invaluable guidance. Every time I speak with him, I am amazed by how brilliant yet how humble he is.

I would like to thank all my friends and lab mates at MIT for their empathy and encouragement through all the crests and troughs of graduate school life. They have made this journey very enjoyable. I would also like to thank all my friends from back home that continue to make me feel valued in their lives.

Finally, I want to thank my family for their unconditional love and support. I thank my mother for encouraging me to choose STEM and for being my best friend. I thank my father for always believing in me and reminding me that I am capable of greatness. And my sister, for inspiring me to be my authentic self.

Being a graduate student at MIT has by far been the most rewarding experience in my life. I continue to learn and be inspired by everyone in this community.

Table of Contents

List of Figures

List of Tables

1. Introduction

The trucking industry in the USA is a major contributor to the nation's economy. Total US business logistics cost accounted for 7.7% of the US GDP in 2017 of which 43% was transportation costs by motor carriers (AT Kearney, 2018). The trucking industry consists of shippers, organizations that have goods which need to be transported; and carriers, organizations that provide the said transportation services. Sometimes a third party logistics provider (3PL) serves as a middle man between shippers and carriers. Ground transportation of freight by motor carriers can further be classified into truckload (TL), less than truckload (LTL), and private fleet. Truckload shipments move from a single origin to a single destination in 48' or 53' trailers. They serve one customer per trip. Less than truckload is for shipments less than 10,000 lbs. where multiple customers can be served in each trip, making multiple stops. In case of TL, the truck may not be physically full but one shipper pays for the entire vehicle-trip. And private fleet is when shippers own their own fleet of trucks. The contribution to the total US business logistics cost in 2017 was $289.4 billion by TL, $62.4 billion by LTL, and $289.6 billion by private fleet (AT Kearney, 2018).

The significance of the industry is also reflected in its size. Trucks move over 70% of freight in the US (Corridore & Chuah, 2018). In addition to being a big market, it is also highly competitive and fragmented. There are more than 1.5 million carrier on record (American Trucking Association. Very few of those carriers are significantly large in size; 91% of the carriers own less than 6 trucks and over 97% have less than 20 (American Trucking Association). Trucking industry is an important and interesting market to be studied. This thesis focuses on the truckload segment of the industry.

Shippers and carriers typically interact through the procurement process as described by Caplice & Sheffi (2005) and Caplice (2007). There are two phases in this process, strategic and operational. The strategic phase starts with shippers determining their projected demands on specific lanes. They send out a call for bidding to a group of carriers that respond with proposed rates for those lanes and specified volumes. The mutually agreed rates are set and are called contract rates. The shipper selects winning carriers for each lane and prepares a routing guide; a list of carriers for a lane in preference order. It is made based on the agreed prices, and is also influenced by the relationship between the two parties and the past performance of the carriers. These contracts are signed for a long term, usually ranging from 1 – 2 years.

The operational phase occurs when a load is ready to be tendered. At that time the shipper offers the load to its primary carrier (the first carrier on the routing guide). If the carrier accepts, the contract rates are paid. But truckload contracts, unlike most other industries, are non-binding in terms of volume. If the carrier doesn't have capacity available or wants to avoid empty backhauls, they can deny service without incurring an explicit penalty in most cases. However, rejecting loads could risk their future business with the shipper. In case of rejection the shipper contacts the next carrier in the routing guide and so on. But prices are usually higher as they go down the routing guide, and the process takes up time and resources. Another option available to the shippers is the spot market. They can hire carriers with available capacity on that lane and pay a one-time price that's decided on a load by load basis. The spot market typically makes up 5-10% of the shipment volume (Caplice, 2007) but increases during a tight market. Spot rates are usually higher and more volatile compared to contract rates and thus more difficult to predict. Some carriers may also reject contract loads expecting to get a better price or volume at the spot market.

The period between the Fall of 2017 and Fall of 2018 was an interesting time for the trucking industry. It witnessed a dramatic growth in demand due to economic development and tax cut benefits leading to more consumer spending, and rise of e-commerce contributing to higher customer expectations. But simultaneously there was a decrease in supply due to a shortage of drivers. The retiring workforce was not being adequately replaced. New laws on increased minimum age of interstate driving forced people, who would have joined trucking right after high school, to opt for other options like construction jobs or college. Additionally, new laws on hours of service and Electronic Logging Device (ELD) mandates contributed to a capacity crunch (Costello, 2017). Carriers shifted their business to lanes with higher returns and re-focused their capacity to the spot market. Furthermore, the US was hit with intense hurricanes during this period. Hurricane Irma and Hurricane Harvey caused capacity to shift to disaster areas leading to scarce services elsewhere. All these forces led to an increase in rates and more frequent reliance on the spot market. (AT Kearney, 2018)

This tightening of the market is not unique. Pickett (2018) describes how the truckload industry goes through cycles of tight and soft market which last around 2 years as shown in Figure 1. A tight market is when demand exceeds supply and the rates are consequently high. A tight market is often called the seller's or the carrier's market, as it is favorable to carriers. New carriers join and existing carriers increase their capacity to capture the demand. When this increase in capacity materializes, the supply increases. When supply exceeds demand, it is a soft market which is favorable to the shippers. New shippers join and existing shippers look to expand. The demand increases again and we get yet another cycle of tight market. Demand spikes are also caused by seasonal holidays and natural disasters. While forecasting short term rates one should keep in mind that such changes occur regularly in the market, and the forecasting models should be able to perform well even as the underlying nature of the market rate changes.

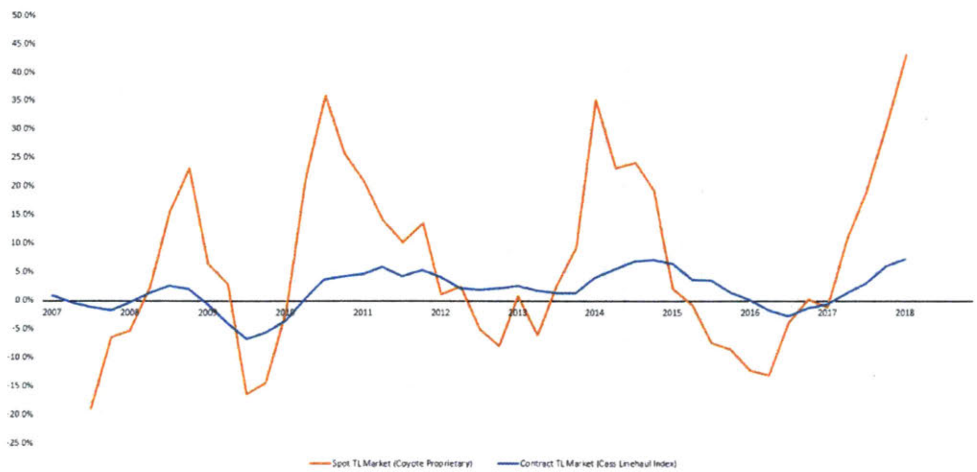


Figure 1. Spot and Contract Rates in Soft and Tight Market Cycles

The objective of this thesis is to understand how rates in the truckload spot market behave and to develop forecasting models to predict them for the near future. The rest of this thesis is organized as follows. In Chapter 2 we motivate the research and outline the research objectives. Chapter 3 reviews literature on freight rate forecasting, forecasting prices using neural networks, and handling concept drift. Chapter 4 discusses the methodology used and in Chapter 5 we look at the results. We finally conclude the thesis in Chapter 6 and discuss future directions of research.

2. Motivation and Research Objectives

Short term spot rate forecasts can help with operational decisions and price negotiations. Carriers can identify high profit and high volume lanes (AT Kearney, 2018) to re-position their resources, and estimate cash flows. Shippers can similarly identify lanes where loads might be rejected (AT Kearney, 2018) and possibly reschedule shipments. Shippers can also use short term spot rate forecasts to decide how far down the routing guide they need to look before opting for spot market, and estimate their operational budget.

Spot rates also influence various contracts. Short term spot rate forecasts act as good indicators for contract rates, capacity, and other market trends in the future (Harding, 2017). Spot rate forecasts can also be used to make decisions in futures market (Asche & Guttormsen, 2002). In March 2019 the industry witnessed the launch of Trucking Freight Futures Contract by Nodal Exchange, FreightWaves, and DAT (HDT Staff, 2019). Futures market is where shippers and carriers choose fixed rates that they are willing to pay and accept, and the investor (a middle man) takes on all the risks and benefits of the actual market rate that materializes. Short term spot rate forecasts can help players identify how much they should bid depending on which direction they want to hedge their risks. Additionally, spot rates are sometimes used to design index based flexible contracts in which the shipper pays a price relative to the market rate (Tsai, Saphores, & Regan, 2011). In such cases too spot rate forecasts can be useful in estimating costs and negotiating prices.

Moreover, transportation costs make up a very significant portion of the total logistics costs for all companies and are used in decision models throughout the supply chain ranging from ordering decisions to facility location planning, transportation mode choice, vehicle routing, and inventory replenishment (Swenseth & Godfrey, 1996). We study short term forecasting of truckload spot rates because it can help all players of the industry in numerous ways.

In this analysis we only look at full truckloads. We focus on long-haul trips ($\geq$ 250 miles) because it has been affected the most by hours of service rules, ELD mandates, and age of interstate driving laws. As long-haul round trips need either multiple days or multiple drivers, the capacity crunch made its effects visible. Unlike contract rates that are usually set months or years in advance and for a particular shipper-carrier pair on any lane remain steady through that period, spot rates are set on a load-to-load basis close to the time of shipment and are more volatile, making it difficult to forecast. Additionally, spot rates only make up a small fraction of the total shipments (~10%) adding the complexity of possible inactivity during certain periods. Aggregating shipments and longer forecast horizons lead to robust future estimates, but the challenge is delivering meaningful forecasts for smaller regions and shorter time horizons.

This thesis aims at answering the following 4 research questions:

1. What are the best models for making short term and frequent forecasts of volatile time series such as truckload spot rates?
2. How do different parameters of the forecasting models, like regional scope, forecasting period and horizon, input variables and their lags, length of training window, and frequency of updating, affect the performance of the model?
3. How should these forecasting models be implemented to handle concept drift?
4. How do these forecasting models perform compared to a simple Naïve model?

We want to see if there is any utility in spending resources into producing a complex forecasting model for short-term forecasting purposes? And if so, what is the best model to choose?

In order to answer these questions, we first study the characteristics of the long-haul truckload market by formulating market indicators based on the observed characteristics of the rates. Then we predict spot rates at a National and 3-Zip origin region level for 1 – 8 weeks into the future using Naïve Models, Moving Average (MA) Models, Auto Regressive Integrated Moving Average (ARIMA) Models, and Feed Forward Neural Network (FFNN) Models. We also use concept drift handling techniques to make better predictions using these forecasting models.

3. Literature Review

In this chapter we review literature on freight rate forecasting for TL, LTL, and ocean freight. We look at usage of regression, auto-regressive integrated moving average (ARIMA), and neural network models, among others, in various cases. Additionally, we also introduce concept drift and discuss methods to handle it. Finally, we outline research gaps that we aim to fill.

3.1 Freight Rate Forecasting

The literature on forecasting freight rates is quite extensive. Many researchers have looked into using regression techniques, especially for forecasting LTL rates. The explanatory variables are usually different combinations of distance, day of the week, economic indicators, and vehicle, load, origin, destination, and market characteristics. Ballou (1991) used a linear regression of distance for LTL rate estimation of different weight breaks. They found that separate rate estimation curves should be modeled for various origin locations for more accuracy. Swenseth & Godfrey (1996) tested five continuous functions to calculate LTL rates for a particular distance. The functions used were constant, proportional, exponential, inverse, and adjusted inverse functions of the shipment weight and corresponding TL rates. The mean squared error of the calculated values were compared to test the performance of the models and determine which one in most useful in which case. The linear function and power function of shipment weight were also used by Mendoza & Ventura (2009) to estimate LTL rates for inbound transportation costs. These functions are recommended by them when there are a large number of suppliers and there is limited access to optimization tools. Kay & Warsing (2009) used multiple non-linear regression to estimate tariff-based rates of LTL as a simple analytical comparison to TL rates for mode choice. They used load density, shipment weight, and origin-destination pair as inputs to the model. multiple non-linear regression was also used by Özkaya, et al. (2010) to predict market rates for

LTL given origin-destination, freight class and weight. To quantify intangible influencers like shipper characteristics and perceived freight rate, they created a Shipper Index and a Freight Index using expert surveys.

Lindsey, et al. (2013) created a regression model to forecast linehaul cost per mile for spot shipments of a US based 3PL company. They used distance, volume-to-capacity ratio, origin and destination characteristics, type of equipment, market indices, and time of the year to determine prices at a lane level and individual shipment level. Scott (2015) modeled load-level forecast of spot premium for a large US based shipper. The regression model took lead time, lane, bid details, calendar week, and carrier into consideration. Their findings show that truckload prices of today influence the prices in the future. More recently Miller (2018) used ARIMA models to make monthly forecasts of Producer Price Index and average spot rates (in Dollars per mile) for full truckloads of dry van and reefers at a national level. These research help understand how different variables influence freight rates in various cases.

Artificial neural networks (ANN) have also been used to predict freight rates using input variables similar to the ones in aforementioned regression techniques. The gain in their popularity is because multilayer feed-forward neural networks (FFNN) have been shown to be able to approximate any continuous function with required accuracy, given adequate network architecture (Hornik, et al., 1989). Li & Parsons (1997) used a FFNN model to forecast ocean tanker freight rates. The results were compared to auto-regressive moving average (ARMA) models. Two types of FFNN models were tested. The first one was auto-regressive and used just monthly tanker spot rates. The second used monthly tanker spot rates, tanker demand, and tanker supply as inputs. FFNN performed better than ARMA, especially for forecasting beyond one month. Lyridis, et al. (2004) used FFNN to forecast monthly spot rates in the Very Large Crude Carrier (VLCC) market for one trade route. Varying numbers of external variables chosen based on correlations were used to forecast rates for different lead times. Variables included demand, production, fleet characteristics, and various market prices. An important observation was that the input variables in difference form performed well during volatile period and rest of the time variables in normal form were better performers. Additionally, the FFNN model outperformed the naïve model for predictions beyond 1 month. Budak, et al. (2017) used FFNN and compared it to a quantile regression model to estimate truckload spot rates for a Turkish logistics company. Input variables related to origin-destination characteristics, distance, load characteristics, vehicle type, prices, and month were used. First the models were run for individual routes in which case FFNN gave better performance than quantile regression. When all the routes were considered together, quantile regression gave better results. Additionally, for both FFNN and quantile regression, route-based model performed better. Most of these cases highlight how ANNs outperform other time series forecasting techniques.

3.2 Forecasting Prices using Neural Networks

Some research has found that ANNs produce better long-term predictions for freight rates as compared to other time series models. Because of their potential, neural network models have also been used to forecast indexes and prices for various markets other than freight rates. Uyar el al. (2016) used fuzzy recurrent neural networks (RNN) trained using genetic algorithm (GA) to predict annual long term freight rate index (LFI) for dry cargo. The model assumed an auto-regressive relationship with a lag of 3 years. They showed that the ANN approach gave better results than other models such as naïve, ARIMA, Holt-Winters and few other similar research proceedings. Sahin, et al. (2018) used a FFNN model to predict weekly Baltic Dry Index (BDI) using past weeks' values and crude oil prices. They comment on the difficulty in forecasting volatile and complex time series data. Zeng, et al. (2016) forecasted daily and weekly BDI using a combined empirical mode decomposition (EMD) and ANN model. Using EMD they first decompose the BDI series into three components and use a FFNN to forecast each of them which are then combined to give a final BDI prediction. Their results show that the combined EMD-ANN model gave better accuracy than ANN or vector auto-regressive models.

Kiani & Kastens (2008) used FFNN and RNNs to forecast foreign exchange rates using autoregressive inputs. They compare the errors of the output to those from an ARMA model to find the most appropriate model for each case. Kristjanpoller & Minutolo (2015) used a hybrid GARCH-ANN model to forecast spot and future Gold price volatility. The FFNN model took the forecast of the generalized autoregressive conditional heteroskedasticity (GARCH) model as inputs along with other relevant financial parameters. The results from the combined model showed 25% less errors than the GARCH model alone. Kölmek & Navruz (2015) used FFNN to predict day-ahead electricity spot price for Turkey. The inputs to the FFNN model included historical values of demand and prices, forecasted demand, production capacity, variables for time and day, and city and contract characteristics. They find that historical values of the time series were the most effective inputs in the model, and the FFNN model performed better than the ARIMA model. Safi & White (2017) predicted daily stock prices of Palestine using FFNN. The FFNN used lagged valued of stock prices as inputs. They too find that the FFNN model performed better than the ARIMA model. Sheta (2018) used a FFNN with autoregressive external inputs to predict daily stock market exchange prices for the Reserve Bank of Australia and compared it to a multiple regression model and report that the FFNN model was superior. This further reinforces that ANNs are a strong forecasting technique and outperform other methods like regression, ARIMA, naïve models etc. in many cases.

3.3 Concept Drift Handling

Concept Drift is the phenomenon when the data evolves over time and the underlying structure and relationships of the dataset changes. In real life data this is a common occurrence. Models built on older datasets become obsolete and need to be updated in order to adapt to these changes. Many researchers like Widmer & Kubat (1996), Tsymbal (2004), and Žliobaitė (2010) discuss methods of handling concept drift. These methods have been applied to multiple studies of classification tasks and time series analysis.

Some researchers use methods that focus on selective weighing of historical data in the model estimation and training. Klinkenberg & Renz (1998) and Klinkenberg (2005) use adaptive selection and filtering for classification of text as relevant and non-relevant. Gu, et al. (2013) improve their forecasts of around 100 different time series using recentness bias learning. Guo, et al. (2018) predict real life and simulated time series with concept drift using adaptive forgetting.

Other researchers adopt combination or ensemble forecasting methods. Kolter & Maloof (2007) test an ensemble method on a synthetic data strem. Zhang, et al. (2016) use a combination forecast for predicting exchange rates. and Cerqueira, et al. (2017) use this method for forecasting various real life volatile time series. Observing the success of concept drift handling in forecasting volatile time series data, we adopt few of the techniques in our work as well.

3.4 Research Gaps

There exists significant literature on freight rate forecasting, but it is very limited for spot rate forecasting of the truckload market. Some of them are discussed earlier like, Lindsey, et al. (2013) and Scott (2015) look into multiple regression models; Budak, et al. (2017) explore FFNN and quantile regression models using multiple input variables; and Miller (2018) studies ARIMA models. But there are a couple of gaps that we would like to address.

All the research mentioned above forecast spot rates or related quantities in monthly buckets for 1 – 6 months into the future. Long term monthly information is valuable for contract rates, but in case of spot rates short term information is more relevant to shippers and carriers as we discussed earlier. Thus, we want to focus on making weekly forecasts for 1 – 8 weeks into the future. Additionally, current research does not talk about updating the prediction models as the market shifts and the underlying data structure changes. We have seen that handling concept drift is a useful tool in time series forecasting, and that truckload market goes through cycles of tight and soft market where the prices may show drastic changes. Thus, we also employ certain concept drift handling techniques in our forecasting models.

4. Methodology

4.1 Data Set

We use shipment transaction details from a leading US based supply chain consultancy company whose clients purchase $68 billion worth of truckload services per year. The details of the dataset are elaborated in Table 1.

Table 1. Details of the Dataset

Field Name	Details
Shipment ID	Each truckload is a unique shipment – 7,248,959 loads
Ship Date	6th April, 2015 – 2nd December, 2018 (191 Monday-Sunday weeks)
Shipper ID	Unique numeric ID – total 17 shippers having shipments in all 191 weeks
Rate Type	Contract or Spot – as tagged by shipper
Mode	Full Truckloads of Dry Van
Origin State	All in the continental USA
Origin Zip	5 digit zip code
Destination State	All in the continental USA
Destination Zip	5 digit zip code
Cost Per Load	Linehaul cost in USD
Distance	In miles ($\geq$ 250 miles)

In our analysis we focus on spot rate information at the national level (all 7.2 million loads) and also at a 3-Zip origin region level. We select the 4 highest volume 3-Zip origin regions, as mapped in Figure 2: Chicago (Zip – 604), Dallas (Zip – 761), Denver (Zip – 804), and Los Angeles (Zip – 917). We select high volume lanes in order to utilize a larger part of the available data set. Additionally, larger volume will ensure a smaller number of weeks with null shipments and thus more complete time series that will be easier to study and forecast.

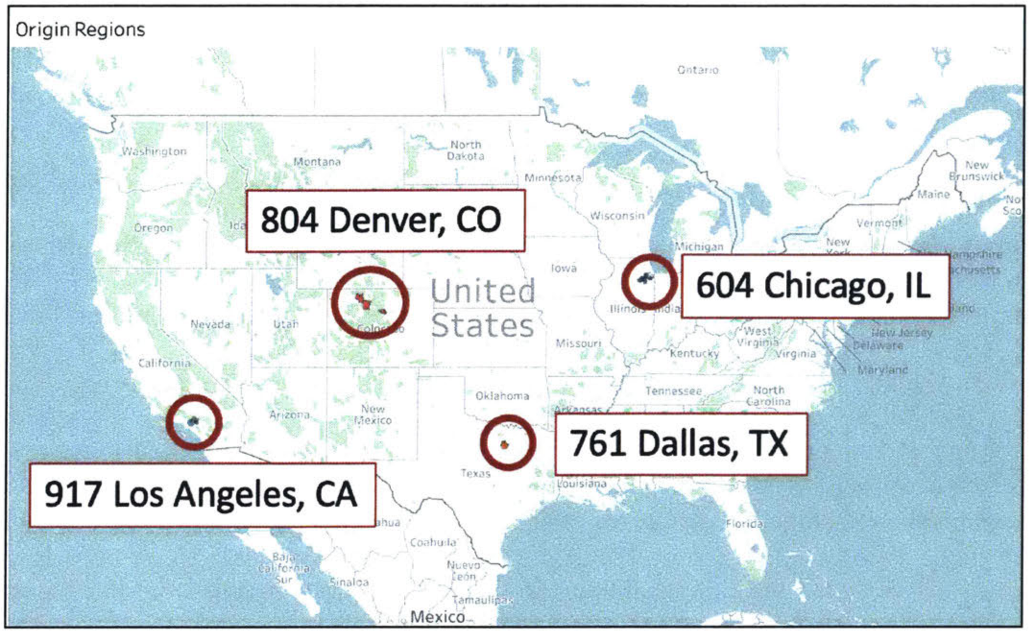


Figure 2. Map of the 4 High Volume 3-Zip Origin Regions Used

4.2 Classification of Rates

Each shipment in the data set is tagged as Contract or Spot by the shipper. We find those tags to be somewhat inconsistent. Figure 3, for example, shows the shipment data for a particular 5Zip-5Zip lane (5 digit zip origin to 5 digit zip destination) from April 2015 to Nov 2018. Each dot on the graph represents a shipment, with the shipment date on the x-axis and cost per mile (CPM) in USD on the y-axis. The steady rates at 3.56, 3.73, and 4.07 are flagged as contract shipments. These tags seem correct, as it is expected that contract rates are stable for a longer term. But we also observe that some shipments tagged as contract do not follow a steady rate, and are instead scattered and look more like spot rates. We speculate that this volatile deviation could be because the shipper was rejected by the primary carrier and had to rely on carriers farther down the routing guide. Another reason could be because of some lack of reliability on how shippers tag the shipments on the data platform. Similarly, we also see that some spot tagged shipments show repeated rates. Moreover, some spot shipments also have lower rates than the average market rates. This could be because the carrier accepted low priced shipments in order to avoid empty trips.

We see that both contract and spot rates exhibit some degree of volatility and there is no explicit information in the dataset that shows the reason behind it. We want to distinguish between the long term stable rates that the shippers plan for and the volatile unexpected rates that they experience. We thus classify the shipment rates based on their frequency of occurrence. For each 3-Zip origin region in our analysis, we look at 5Zip-5Zip lanes that have at least 1 load per month. We choose lanes with higher volumes for easier classification and larger amount of information. For each lane we classify the CPM as Recurrent Rate (R) if it has appeared 1 or more times in the previous 4 weeks, and as New Rate (N) if it has never been seen in the previous 4 weeks. Rates that haven't been seen in recent past are a proxy for spot-like behavior and reappearing rates are a close substitute for contract-like behavior. The Recurrent and New tags give us more information than just the shipper identified Contract and Spot tags.

The decision for using a 4 weeks window for classification is not picked optimally. A smaller window size, like 1 week, would cause more shipments to be tagged as N. This could result in some regular rates that miss a few weeks to also be considered as spot-like activity. Whereas a longer window, like 12 weeks, will result in more shipments to be tagged as R. The number of shipments tagged as R or N changes in each case but we believe that the trends over time will follow similar patterns. Other window sizes may be tried and tested, or even optimally chosen in future extensions.

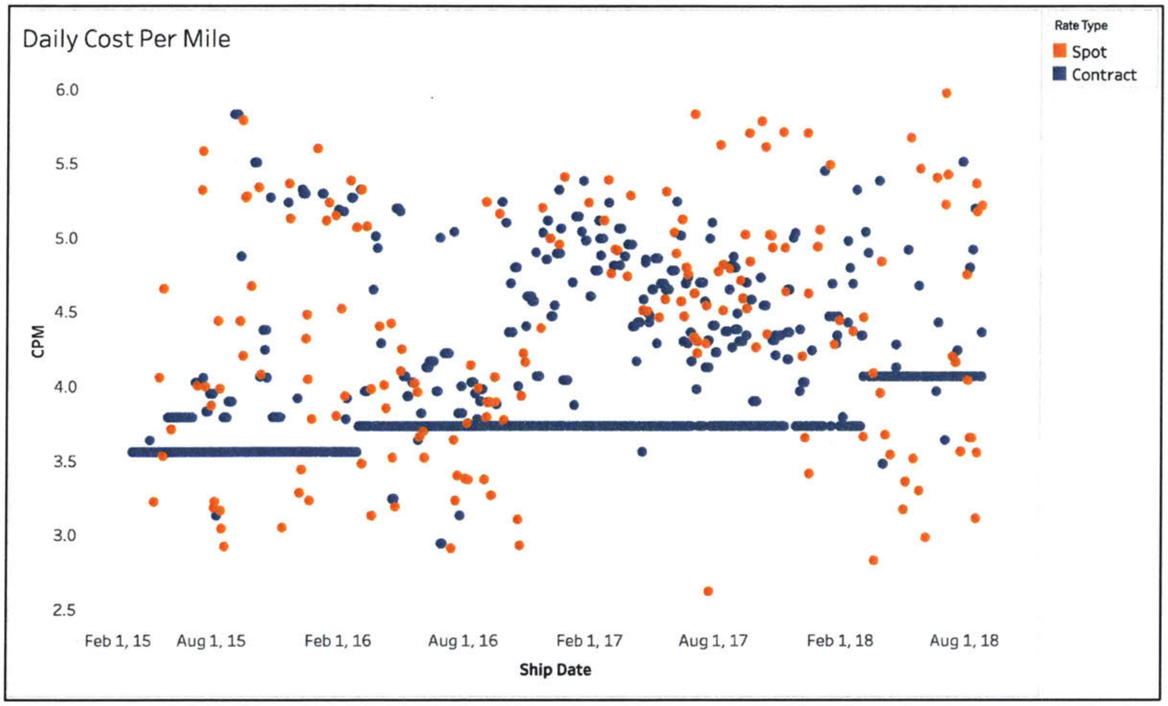


Figure 3. Shipment Data of a 5Zip-5Zip Lane

In Table 2 we observe some details of the origin regions whose shipments are classified. We see that Dallas has the lowest number of lanes and Denver has the lowest number of shippers of all the regions. This could cause some volatility due to lack of aggregation. On the other hand, Chicago and Los Angeles have larger number of shippers and lanes. Aggregation in these regions may exhibit more stability in values of rates. We want our prediction models to be able to forecast both kinds of time series adequately.

Table 2. Origin Regions Details

3Zip Origin Region	Average Annual Volume	# of Shippers	# of Lanes
604 (Chicago, IL)	16,133	8	94
761 (Dallas, TX)	3,096	4	12
804 (Denver, CO)	44,336	1	169
917 (Los Angeles, CA)	43,560	6	167

4.3 Formulation of Indicators

After classifying the rates in each lane l in week w, we will observe one of the following cases as shown in Table 3:

1. Neither spot-like nor contract-like activity present i.e. no shipments
2. Spot-like activity present but no contract-like activity
3. Contract-like activity present but no spot-like activity
4. Both spot-like and contract-like activity present

Table 3. Cases for Presence of Spot and Contract Like Activity

		No New Tags	New Tags
		$\boldsymbol{n_{N,l,w} = 0}$	$\boldsymbol{n_{N,l,w} > 0}$
No Recurrent Tags	$\boldsymbol{n_{R,l,w} = 0}$	Case 1	Case 2
Recurrent Tag	$\boldsymbol{n_{R,l,w} > 0}$	Case 3	Case 4

In order to characterize the spot market, the indicators should reflect the status of spot-like activity for each of these cases. We identify four measures, that are useful in understanding the characteristics of the truckload market, as enumerated in Table 4. The indicators at the national and 3-Zip origin level are similar but not identical. At the national level we calculate the Contract and Spot Rate Index as the average Contact or Spot rate in a week, indexed off the base week. At the national level we simply use the shipper specified Contract and Spot tags because classifying more than 7 million loads on tens of thousands of 5Zip-5Zip lanes is computationally intensive. We also calculate the Spot Premium Ratio that is the ratio of average Spot rates to average Contract rates in a particular week. And the Spot Volume ratio that shows what percentage of the total shipments in a week were under Spot. Similarly, for the 3-Zip origin region level we calculate Recurrent and New Rate Index as the average Recurrent or New rate in a week, indexed off the base week. For the New Premium Ratio, we take the ratio of average New rates to average

Recurrent rates of each 5Zip-5Zip lane starting from the origin region, and calculate the volume weighted average of these values for each week. And finally, the New Volume ratio is simply the ratio of number of shipments tagged as New to the total number of shipments in a week.

Table 4. Definition and Formulation of Market Indicators

Level	Indicator	Definition	Formulation
National	Contract Rate Index	Average contract rates for all shipments in week w indexed off the starting week of the data set	$CRI_w = \frac{r_{C,w} \times 100}{r_{C,base}}$
	Spot Rate Index	Average spot rates for all shipments in week w indexed off the starting week of the data set	$SRI_w = \frac{r_{S,w} \times 100}{r_{S,base}}$
	Spot Premium Ratio	Ratio of average spot rates to average contract rates in week w	$SPR_w = \frac{r_{S,w}}{r_{C,w}}$
	Spot Volume Ratio	Ratio of number of all spot rate loads to total loads in week w	$SVR_w = \frac{n_{S,w}}{n_w}$
Regional	Recurrent Rate Index	Average recurrent rates for all shipments in week w indexed off the starting week of the data set	$RRI_w = \frac{r_{R,w} \times 100}{r_{R,base}}$
	New Rate Index	Average new rates for all shipments in week w indexed off the starting week of the data set	$NRI_w = \frac{r_{N,w} \times 100}{r_{N,base}}$
	New Premium Ratio	Volume weighted average of ratio of average new rates to average recurrent rates in week w for all lanes l of an origin region	$NPR_{l,w} = \frac{r_{N,l,w}}{r_{R,l,w}}$ $NPR_w = \frac{\sum_l n_{l,w} NPR_{l,w}}{\sum_l n_{l,w}}$
	New Volume Ratio	Ratio of number of all new rate loads to total loads in week w	$NVR_w = \frac{n_{N,w}}{n_w}$

R = Recurrent Rate (classified by us)
N = New Rate (classified by us)
C = Contract Rate (tagged by shipper)
S = Spot Rate (tagged by shipper)
l = 5zip-5zip lane
w = week
n_x = number of loads of rate type x
r_x = average Cost Per Mile (CPM) of rate type x

4.4 Short Term Forecasting

The objective of this research is to create useful short term forecasts of the truckload spot market. We identify Spot/New Rate Index and Spot/New Premium Ratio as the potential metrics to forecast as both give information about behavior of spot rates in the TL market. To select the indicator to forecast we perform tests to check for Random Walk. A Random Walk is a stochastic process in which the value in a sequence is a random step from the previous values in the sequence. This implies that to predict a Random Walk the most appropriate model is a naïve model, in which the forecast of the next time step is the value at the current time step. In such a case testing other forecasting models would be unnecessary. To test for Random Walk, we check for the following 3 properties in the time series:

1. In Random Walk processes, auto-correlation shows strong positive values for smaller lags and then decays linearly.

2. Random Walk time series are non-stationary i.e. the properties of the series like mean, variance and covariance change over time (Montgomery, et al., 2016). We check for stationarity using the ADF test.

3. Making a Random Walk time series stationary by differencing shows no observable learnable structure. The autocorrelations of the differenced series are close to 0 for all lags.

The results of these tests are summarized in Table 5. We choose to forecast Spot/New Premium Ratio because they are more likely to not be Random Walks and thus various forecasting models can be tested. Additionally, Spot/New Premium Ratio provides vital information on what the spot rates will be in comparison to contract rates. Buyers and sellers of transportation services make decisions whether to opt for spot market or not based on such information. Spot/New Premium Ratio is also an indicator for the direction of change in the market.

Table 5. Results of Tests for Random Walk

Region	Spot/New Rate Index			Spot/New Premium Ratio		
Property	**1**	**2**	**3**	**1**	**2**	**3**
National	Yes	Yes	Yes	Yes	Yes	Yes
Chicago (604)	Yes	Yes	Yes	No	No	Yes
Dallas (761)	No	Yes	Yes	No	No	Yes
Denver (804)	No	No	Yes	No	No	Yes
Los Angeles (917)	Yes	Yes	Yes	No	No	Yes

We predict weekly Spot Premium Ratio at the national level and New Premium Ratio at the 3-Zip origin region level. We make forecasts every week for 8 weeks into the future (forecast horizon). Transportation activities are usually measured in monthly buckets, while economic activities in weekly buckets. One week is not enough time react to new information when one may need to re-position resources. Whereas planning for operations is not done for beyond 4 – 8 weeks. Thus, we choose to forecast for 8 weeks into the future inspired by practices followed in the industry. The forecasting models are trained on the data set for 2015-2016 and predictions are made for 2017-2018. The 4 models used are defines as follows.

4.4.1 Naïve Model

Naïve models are useful for formulating base case scenarios to compare performance of other sophisticated models to. In cases like ours however, where we try to forecast volatile time series with the possibility of them being a Random Walk, Naïve models can be helpful due to their highly responsive nature. We forecast SPR/NPR for weeks $w+1$ to $w+8$ to be the value of SPR/NPR at week w.

$$\widehat{SPR}_{w+1} = \widehat{SPR}_{w+2} = \widehat{SPR}_{w+3} = \widehat{SPR}_{w+4} = \widehat{SPR}_{w+5} = \widehat{SPR}_{w+6} = \widehat{SPR}_{w+7} = \widehat{SPR}_{w+8} = SPR_w$$

4.4.2 Moving Average (MA) Model

We use another fundamental and responsive time series forecasting method, Moving Average (Montgomery, et al., 2016). MA models provide more stable forecasts than Naïve models as they consider an average of larger amount of historical information, in our case, data from the previous 4 weeks. We forecast SPR/NPR for weeks $w+1$ to $w+8$ to be the average of values of SPR/NPR in weeks $w-3$ to w.

$$\widehat{SPR}_{w+1} = \widehat{SPR}_{w+2} = \widehat{SPR}_{w+3} = \widehat{SPR}_{w+4} = \widehat{SPR}_{w+5} = \widehat{SPR}_{w+6} = \widehat{SPR}_{w+7} = \widehat{SPR}_{w+8}$$
$$= avg\ (SPR_{w-3},\ \ SPR_{w-2},\ SPR_{w-1},\ SPR_{w})$$

4.4.3 Auto-Regressive Integrated Moving Average (ARIMA) Model

ARIMA models (Montgomery, et al., 2016) are another popular method of forecasting time series. The auto-regressive part (p = lags of auto-regression) accounts for the dependence of the forecast of the next time step on the lagged values in the time series. The integrated part is the degree of differencing (d) performed on the series to make it stationary. Differencing of the first order implies we subtract the observation at each time step from the observation of the previous time step to create a new time series that is stationary. And the moving average part looks into the influence of the errors of the moving average forecast (q = size of moving average window) on the prediction of the next time step.

For various values of $p, d,$ and q, ARIMA models can replicate Naïve, MA and exponential smoothing models . We explore all the mentioned options by finding the best fit ARIMA (p, d, q) model for each region using the pyramid-ARIMA package in Python. The function searches through all specified values of the parameters and chooses the one with the lowest AIC. We set the parameters of the model to be in the following ranges: $p \in [1,24]$ i.e. the auto-lags vary from previous 1-24 weeks; $d \in [0,2]$ i.e. the time series is differenced by 0-2 degrees; and $q \in [0,4]$ i.e. the error of the moving average of previous 0-4 weeks influences the future forecast.

4.4.4 Feed-Forward Neural Network (FFNN) Model

We discussed earlier how multilayer feedforward neural networks have been shown to be able to approximate any continuous function with required accuracy, given adequate network architecture (Hornik, et al., 1989). Given the sparse and volatile nature of spot rates and the consequent difficulty in forecasting them, FFNN models can prove to be useful in finding a relationship between specified inputs and outputs if one does exist. We model the FFNNs (Montgomery, et al., 2016) in MATLAB with varying architecture and pick the one with least complexity which still gives comparable results.

The network used has 4 input neurons, 1 hidden layer with 4 neurons, and 1 output neuron as illustrated in Figure 4 (output from executed code). The inputs to the model are lags of $CRI/RRI, SRI/NRI,$ $SPR/NPR,$ and SVR/NVR ranging from the previous $1-4$ weeks to values from 6 months prior. For example, to predict values of SPR_{w+1} to SPR_{w+8} at week w using previous SPR values as inputs, we use the following 6 group of lags as inputs:

1. $SPR_{w-3},\ SPR_{w-2},\ SPR_{w-1},\ SPR_{w}$
2. $SPR_{w-7},\ SPR_{w-6},\ SPR_{w-5},\ SPR_{w-4}$
3. $SPR_{w-11},\ SPR_{w-10},\ SPR_{w-9},\ SPR_{w-8}$
4. $SPR_{w-15},\ SPR_{w-14},\ SPR_{w-13},\ SPR_{w-12}$
5. $SPR_{w-19},\ SPR_{w-18},\ SPR_{w-17},\ SPR_{w-16}$
6. $SPR_{w-23},\ SPR_{w-22},\ SPR_{w-21},\ SPR_{w-20}$

The transfer function for the first layer is a log-sigmoid function: $logsig(x) = \frac{1}{1+e^{-x}}$, and for the second layer we have a linear transfer function. More discussion on the network architecture is done in Appendix A.

The data is first pre-processed using the 'mapstd' function to map the mean of the data set to 0 and standard deviation to 1. We then split the data using the 'dividerand' function into 70%

training set, 15% validation set, and 15% test in a non-chronological random order. We use the 'fitnet function' and train the model using the Levenberg-Marquardt backpropagation training algorithm. It is the fastest and most appropriate algorithm for function fitting tasks for smaller networks (up to few 100 weights) like ours. The weights and biases of the network are set to minimize errors on the training set. The validation set stops the training early if the performance on the validation set doesn't improve or remains constant. And the test set doesn't affect the training process. It is used as an unbiased measure of the network's performance. In the Levenberg-Marquardt backpropagation training algorithm, the training stops when any of the pre-specifies conditions on number of iterations, run-time, performance goal, or minimum gradient occurs.

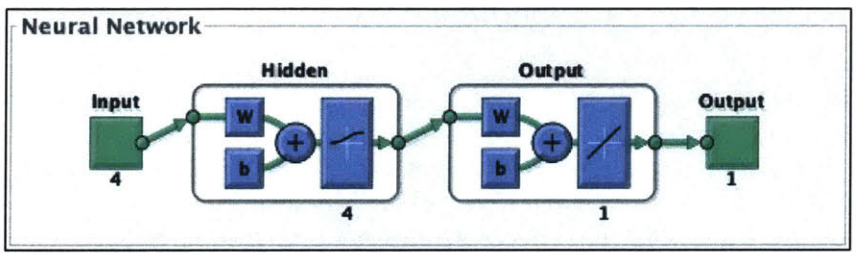


Figure 4. Neural Network Architecture

4.4.5 Handling Concept Drift

Concept drift occurs when the relationships in the data change over time. It is necessary to update our forecasting models in such cases because the model trained on the older data structure is no longer relevant to make useful predictions based on current observations. In our dataset we observe a dramatic increase in spot rates starting Fall of 2017. We want our models to perform well even when market changes from soft to tight or vice-versa every 2 years or so. A few concept drift handling methods described by Žliobaitė (2010) are used here to update the ARIMA and FFNN forecasting models.

We perform 3 types of training for the ARIMA and FFNN forecasting models as illustrated in Figure 5:

1. Single Training – The models are trained only once and never updated. This serves as a base case scenario. This case considers that the underlying data structure doesn't change and thus updating forecasting models is not required.

2. Expanding Window – We update the existing models every 4, 8 or 12 weeks to adapt to the patterns in the new observations. In this case the full history of the time series in considered relevant in the model.

3. Rolling Window – We re-train the model from scratch every 4, 8 or 12 weeks using data from only the previous 52 weeks of data. In this case the older data set is no longer considered relevant in predicting current values.

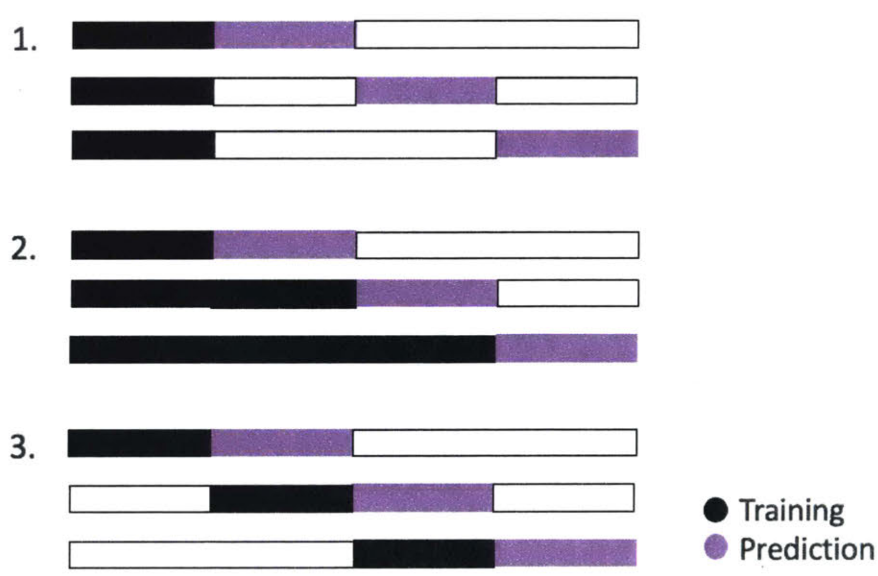


Figure 5. Different Types of Training Scenarios to Handle Concept Drift

5. Results

5.1 Classification

We classify the rates of the shipments based on their frequency of occurrence, according to the rules stated previously. We observe the distribution of the 4 cases in the 3-Zip origin regions over the period of April 2015 to November 2018 in Figure 6. For 5zip-5zip lanes in all 4 origin regions we observe that in most cases there is only contract-like activity ($n_{R,l,w} > 0$) and no spot-like activity ($n_{N,l,w} = 0$) present. Vast majority of lanes have only contract or recurring rates each week and very little of spot or new rates. This is in line with the fact that approximately 90-95% of the truckload shipments are contract rates.

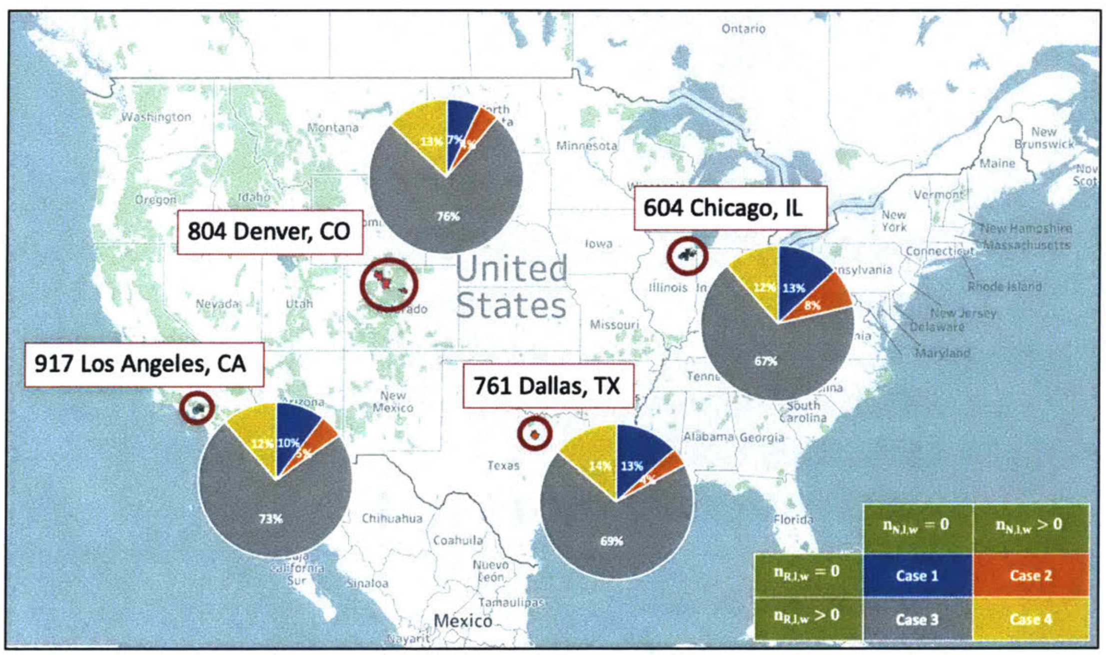


Figure 6. Distribution of Occurrences of Classification Cases for 3-Zip Origin Regions

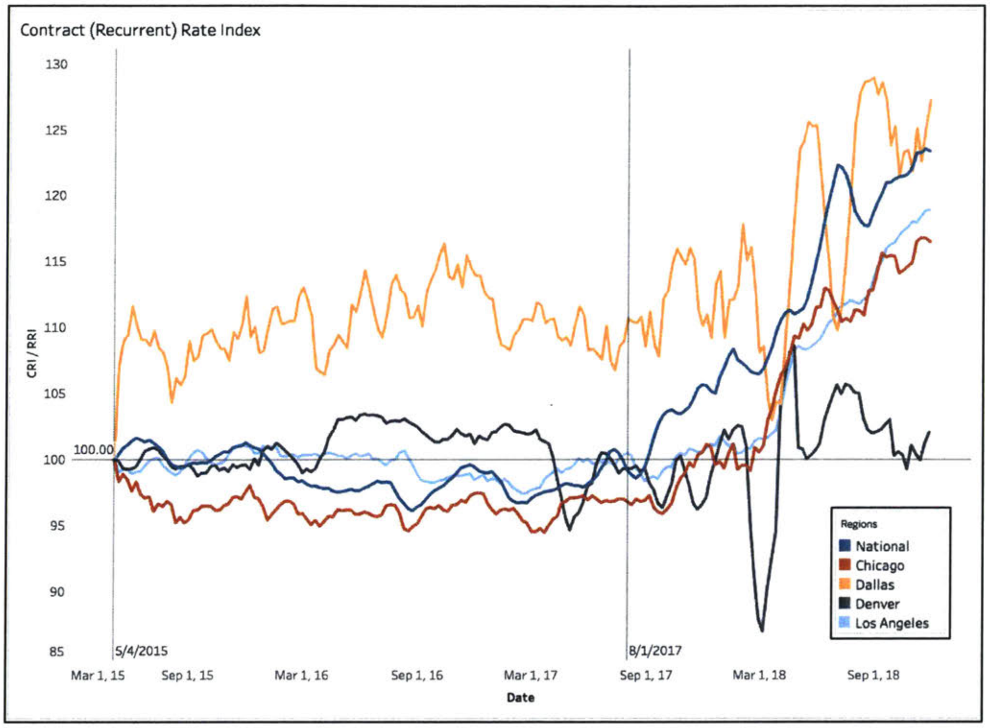
5.2 Trends of Indicators

After classification of the rates, we calculate the values of the 4 aforementioned indicators for the national level and the 3-Zip origin region level. The trends of these indicators are shown in

Figure 7. One major observation is that all indicators increase in value after Fall of 2017 (marked by Aug 1 in graphs). This is consistent with the tightening of the market that happened in that period leading to an increase in rates and shifting to the spot market. The spot rates, premium of spot rates over contract rates, and number of spot loads all increased.

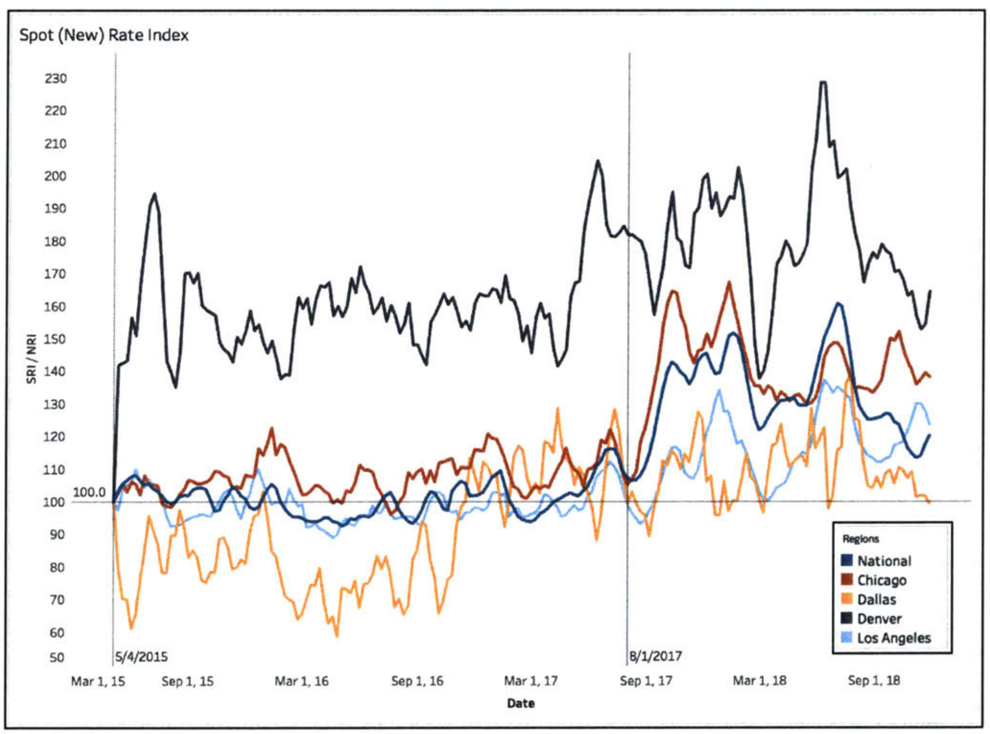
Figure 7(a) shows the Contract/Recurrent Rate Index. The values are fairly stable before Fall of 2017 after which they increase. For Denver we see that there is no prominent increase in RRI. This could be because the region has only 1 shipper in the dataset and that shipper may not have changed their behavior as the market changed. We also see that Dallas shows higher values and more volatility than others. In Figure 7(b) we observe that the increase in Spot/New Rates was greater than the increase in Contract/Recurrent Rates. Spot/New Rates are also more volatile than Contract/Recurrent Rates.

In Figure 7(c) we observe that values for Spot/New Premium Ratio are almost always greater than 1. Spot rates are usually higher than contract rates because they are set on a load-by-load basis closer to the time of shipment, and that is reflected in our dataset. And for regional level we can interpret it as the incoming rates being higher than the older observed rates, indicating that the market rate is slowly going up. Again, we observe the high volatility in Dallas. Lastly, the increase in Spot/New Volume Ratio in Figure 7(d) indicates the shifting of shippers and carrier to spot market in this period. Number of Spot/New loads doubled from being an average of 5% of the total loads to an average of 10% of the total loads. The high peaks could be because of new rates coming in which are tagged as New but later get classified as Recurrent.



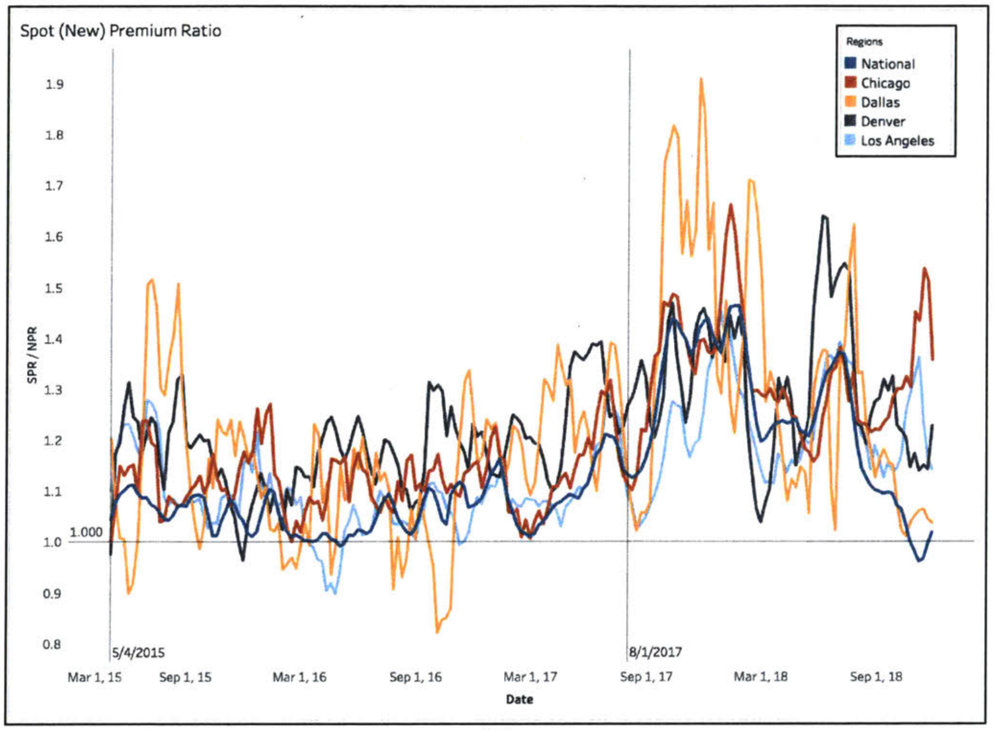
Contract (Recurrent) Rate Index
130
125
120
115
110
105
100
95
90
85
CRI / RRI
100.00
5/4/2015
8/1/2017
Mar 1, 15
Sep 1, 15
Mar 1, 16
Sep 1, 16
Mar 1, 17
Sep 1, 17
Mar 1, 18
Sep 1, 18
Date
Regions
National
Chicago
Dallas
Denver
Los Angeles

(a)

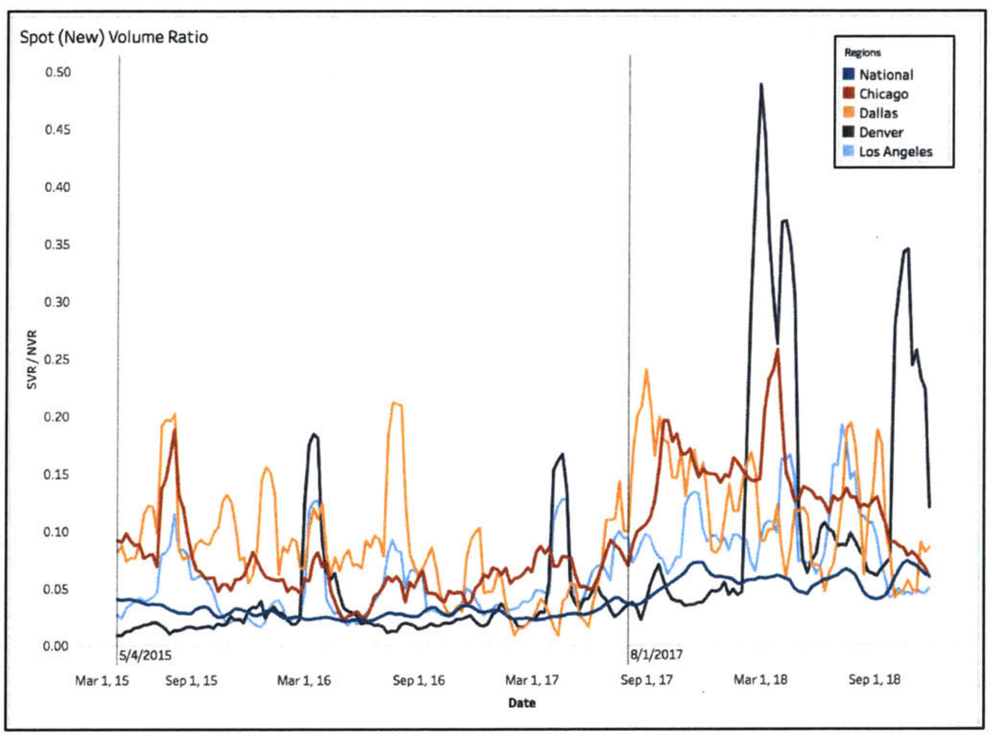


Spot (New) Rate Index
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
SRI / NRI
100.0
5/4/2015
8/1/2017
Mar 1, 15
Sep 1, 15
Mar 1, 16
Sep 1, 16
Mar 1, 17
Sep 1, 17
Mar 1, 18
Sep 1, 18
Date
Regions
National
Chicago
Dallas
Denver
Los Angeles

(b)



(c)



(d)

Figure 7. 4 Weeks Moving Average of Indicators (a)Contract/Recurrent Rate Index (b) Spot/New Rate Index (c)Spot/New Premium Ratio (d)Spot/New Volume Ratio

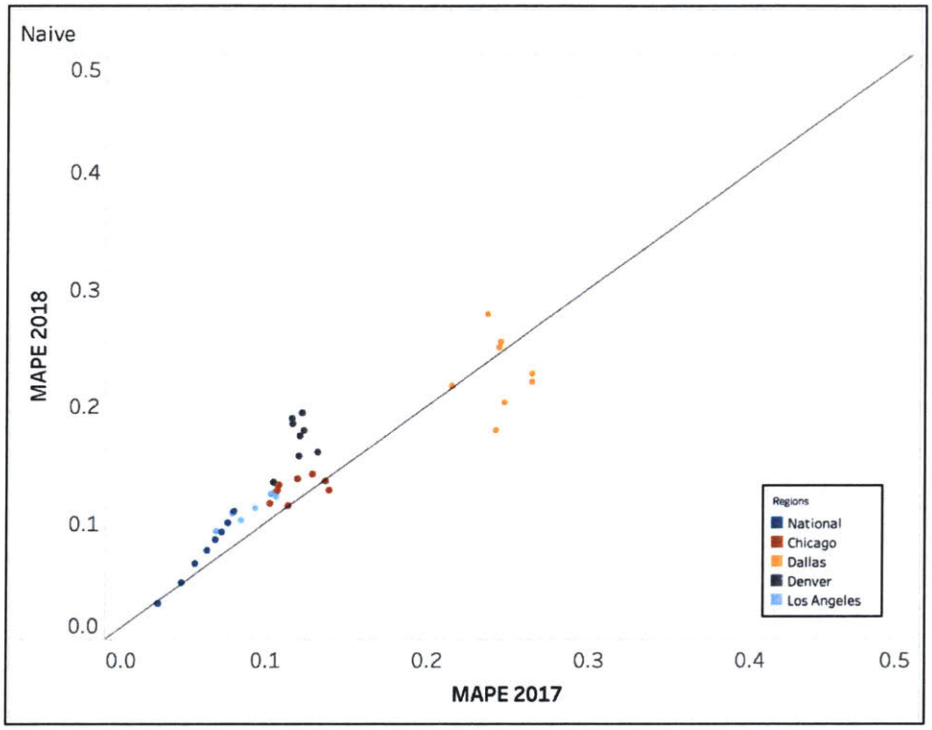
All indicators show trends that are coherent with what the industry experts claimed and stakeholders reported post the Fall of 2017 and otherwise. This demonstrates that the dataset is a good representation of the overall truckload industry. Additionally, we can safely say that the computed indicators adequately capture the characteristics of the long haul truckload market.

5.3 Forecasting Results

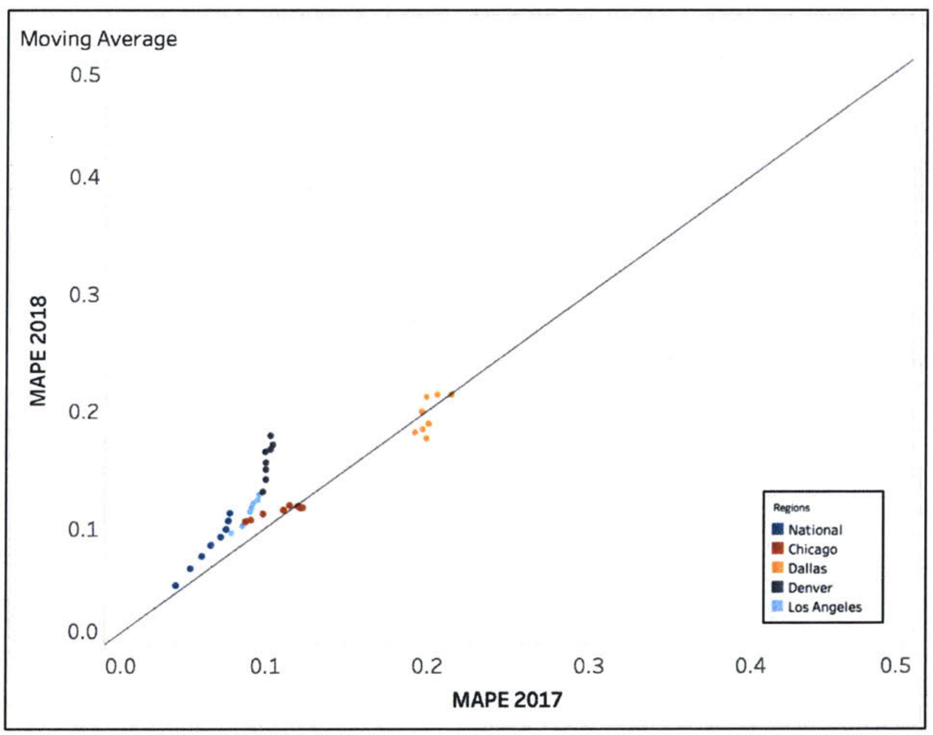
We forecast the national SPR and the regional NPR using the 4 forecasting models and the 3 training methods. We calculate the mean absolute percentage error (MAPE) of predictions made in 2017 and 2018 for all forecasting horizons i.e. predictions for weeks $w + 1$ to $w + 8$, at each week w. We use MAPE as our performance metric because it is scale independent which helps us compare results of different regions. It is also widely used and easy to understand. The results of the forecasts are illustrated in Figure 8. Each point on the graph represents the output of a forecasting model for all combinations of input variable, lag of input variable, forecast horizon, type of training, and frequency of updating.

One of our research questions was to summarize how different parameters of the forecasting models, like regional scope, forecasting period and horizon, input variables and their lags, length of training window, frequency of updating etc., affect the performance of the model. Some overall key insights from the prediction results of the models to answer that question are enumerated below:

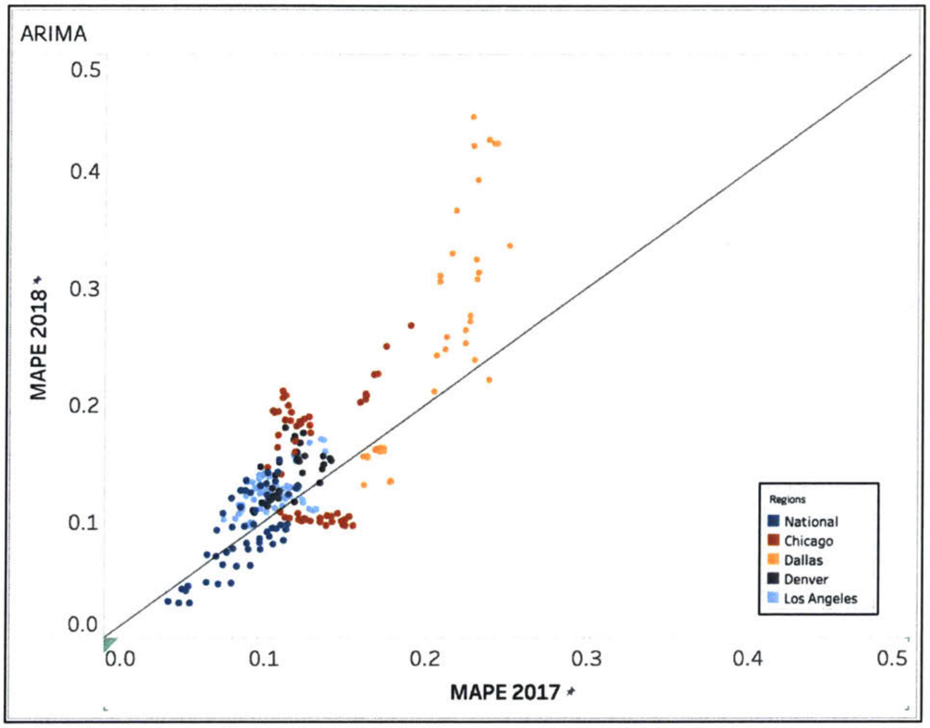
1. Regional Scope: We observe that MAPEs of all models within a region form clusters. This implies that different models predicting a region's SPR/NPR values have comparable performance as opposed to a single model producing comparable results for all regions. This can also be observed in Table 6. The lowest MAPE for each model in each region is comparable for


Naive
0.5
0.4
0.3
0.2
0.1
0.0
MAPE 2018
0.0
0.1
0.2
0.3
0.4
0.5
MAPE 2017
Regions
National
Chicago
Dallas
Denver
Los Angeles

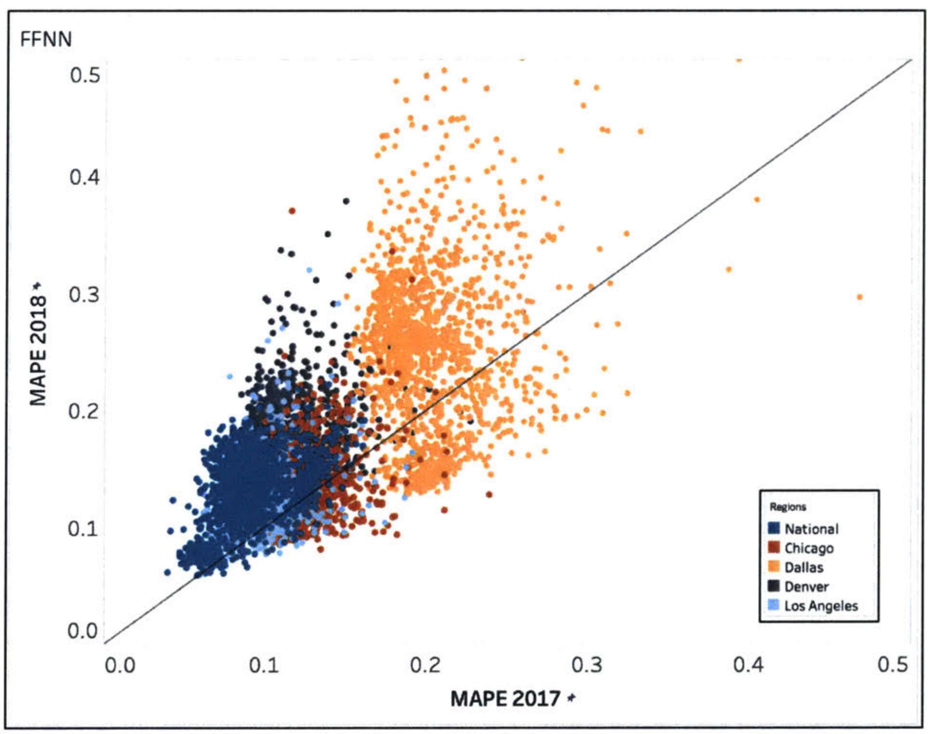
(a)


Moving Average
0.5
0.4
0.3
0.2
0.1
0.0
MAPE 2018
0.0
0.1
0.2
0.3
0.4
0.5
MAPE 2017
Regions
National
Chicago
Dallas
Denver
Los Angeles

(b)



(c)



(d)

Figure 8. MAPE of Forecasting Models (a)Naïve (b)MA (c)ARIMA (d)FFNN

most cases. We also see that MAPEs for Dallas are the highest and that for National level prediction are the lowest among all regions. This is because the values of SPR for the national level are more stable compared to other regions and thus easier to predict. Additionally, the higher MAPEs in Dallas can be attributed to higher volatility of the regions NPR as we have seen before.

Table 6. MAPEs of Best Fit Models

Region	Naïve	MA	ARIMA	FFNN
National	0.069	0.075	0.076	0.067
Chicago (604)	0.124	0.111	0.116	0.112
Dallas (761)	0.236	0.198	0.154	0.168
Denver (804)	0.145	0.129	0.110	0.120
Los Angeles (917)	0.104	0.099	0.095	0.093

2. Forecast Period: In Figure 8 we also observe that in most cases, MAPEs for 2018 are higher than MAPEs for 2017. The first half of 2017 followed trends similar to 2015-2016 and thus the models trained on period 2015-2016 made more accurate forecasts for half of 2017 as compared to 2018.

3. Forecast Horizon: As expected, MAPEs usually increase with forecast horizon. MAPEs of predictions for 1 week out are lower than 2 weeks out and so on, keeping the other parameters constant in a model. We can observe this in the example presented in Figure 9. The chart shows the MAPEs for naïve forecast for national SPR for different forecast horizons. This is as one would expect. But MAPE for all 8 predictions are in the same order of magnitude. Best fit models for each region for each forecast horizon are discussed in Appendix B.

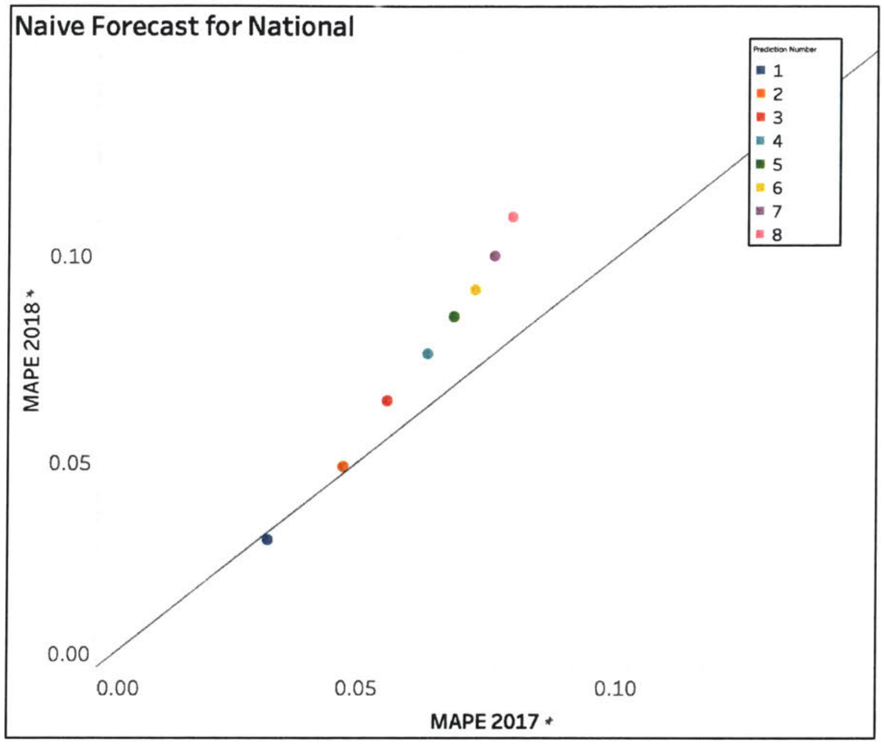


Figure 9. MAPE for Different Forecast Horizons

4. <u>Input Variables and Lags</u>: Input variables and lags of input variables in FFNN models don't have consistent effects across all regions. We can see that in Figure 8(d), the MAPEs are scattered in a wide range for different combination of parameters in the FFNN model. The parameters of the best fit model of each region are given in Table 7. For Chicago, Dallas, and Denver the best fit models use auto-lags of previous 1 – 4 weeks as inputs. In case of National and Los Angeles predictions, the best fit model is a FFNN model which uses SVR/NVR as inputs.

We check to see if SPR/NPR work as good inputs for FFNN in all regions. We compare the MAPE of the best fit FFNN model of each region to the FFNN models that takes the previous 1 – 4 weeks of SPR/NPR as inputs. The results in Table 8 show that the models are comparable. Additionally, the best ARIMA model for all the regions used an auto-lag of 1, except for Chicago which used 9. Thus, we can conclude that in all cases, auto-lags of previous 1 – 4 weeks can be used as inputs to a model to produce comparable results to the best fit models.

Table 7. Parameters of Best Fit Models

Region	Model	Training Type	Parameters	MAPE
National	FFNN	Expanding Window – Update every 4 weeks	Input – SVR Input Lag – previous 13-16 weeks	0.067
Chicago (604)	MA	-	-	0.111
Dallas (761)	ARIMA	Single Training	ARIMA (1, 0, 0)	0.154
Denver (804)	ARIMA	Expanding Window – Update every 12 weeks	ARIMA (1, 0, 0)	0.110
Los Angeles (917)	FFNN	Rolling Window – Update every 4 weeks	Input – SVR Input Lag – previous 13-16 weeks	0.093

Table 8. Comparison of MAPEs of FFNN Models

Region	MAPE for Best Fit FFNN Model of Each Region	MAPE When Previous 1 – 4 Weeks of SPR/NPR are Used as Inputs in FFNN Model for Each Region
National	0.067	0.071
Chicago (604)	0.112	0.119
Dallas (761)	0.168	0.170
Denver (804)	0.120	0.123
Los Angeles (917)	0.093	0.095

5. Length of Training Window: In most cases, training by the Expanding Window method gave lower MAPEs than Rolling Window training, keeping other parameters constant. Models trained on the full dataset of historical values performed better than those that neglect older data in the training. This implies that all of the history of the time series provides valuable information on short-term dependence in the volatility of the values of SPR/NPR.

6. Updating Frequency: In most cases, MAPE for models that were updated every 4 weeks was lower than those updated every 8 weeks which were in turn lower than those updated every 12

weeks, keeping other parameters constant. Updating the models more frequently lowers the MAPE as expected.

Another research question was to find how to implement the models to handle concept drift. Figure 10 shows the forecasts of national SPR using FFNN models using the 3 training methods as compared to the actual values of the national SPR. The FFNN models used take in auto-lags of previous 1 – 4 weeks as inputs, and for the Expanding and Rolling Window method we update the models every 4 weeks. In the graph we observe that the models trained using the Expanding Window and Rolling Window method produce forecasts that closely follow the overall increasing and decreasing trends of the actual values. But in case of the model trained using the Single Training method, the forecasts do not follow the trends of the actual series and only

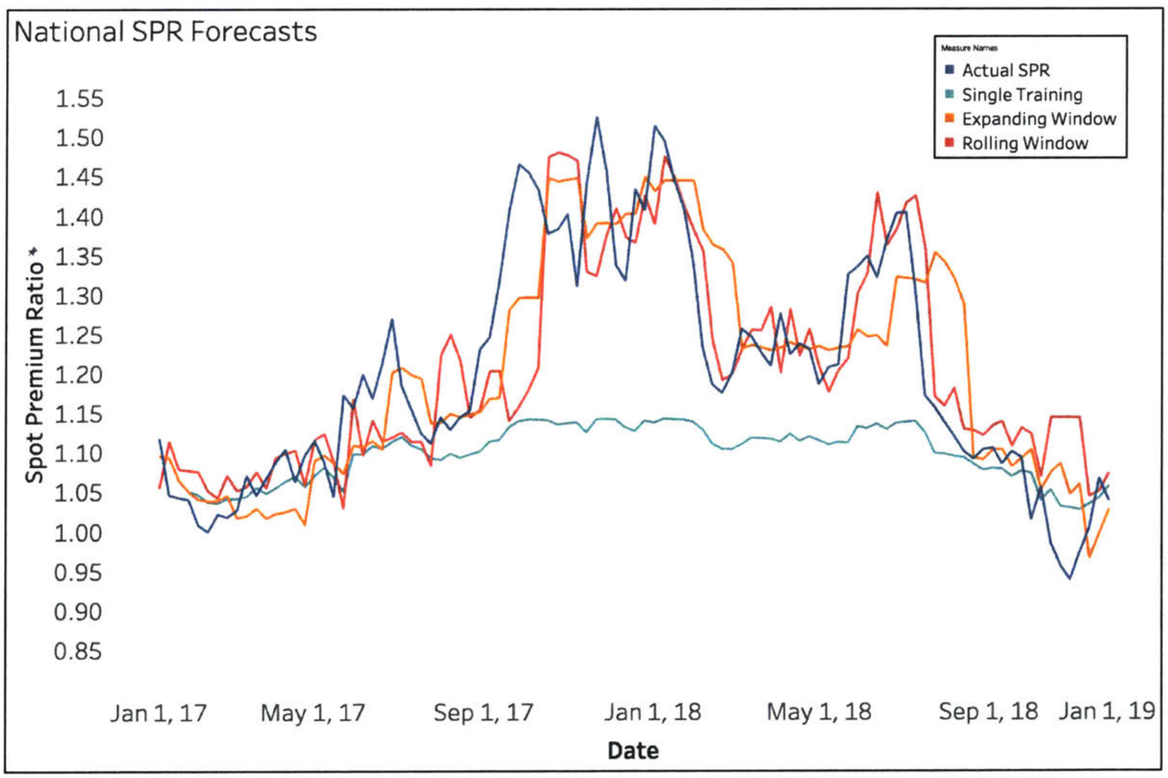


Figure 10. Comparison of Forecasts of FFNN Using Different Training Methods

fluctuate slightly around a stable value. The model was trained on a period with stable rates and produced outputs in the same range of values even when the market changed. This is because the bias term was dominant in the neural network and the inputs only contributed to fluctuations around that constant. This illustrates the need for handling concept drift by regularly updating the forecasting models.

We also want to answer how much better do the forecasting models perform as compared to a simple naïve model for making short term and frequent forecasts of volatile time series such as truckload spot rates. The naïve model for National SPR gives second lowest error (6.9%) amongst all models in all regions. Additionally, it is also easier to implement as compared to the FFNN model for National SPR that gave the least error (6.7%). We test if the naïve forecasts of the National SPR can be used as forecasts for NPR of the 3-Zip origin regions. In table 9 we compare the MAPE of the best fit model of each region with the model that uses naïve forecast of National SPR for all regions. We also compare naïve forecast of individual regions for all forecasting horizons to the case using naïve forecast of National SPR for all regions, in Appendix C. In Table 9 we observe that for most cases the difference in the two MAPEs is negligible, between 0.2-2%.This implies that naïve forecasts for National SPR can very well be used as forecasts for the national level and for most 3-zip origin regions.

Lindsey, et al. (2013) used a regression model to forecast linehaul cost per mile for spot shipments of a US based 3PL company to determine prices at a lane level and individual shipment level. They find that the predictions were within $21-25\%$ of the actual values. Budak, et al. (2017) used a FFNN model along with a quantile regression model to estimate truckload spot rates for a Turkish logistics company. For the route based approach the FFNN model produced a MAPE of 0.8% and quantile regression gave a MAPE of 1.37%. And for the aggregate general approach

FFNN model produced a MAPE of 9.4% and quantile regression gave a MAPE of 6.7%. Whereas the MAPE for using naïve forecasts of National *SPR* in our case ranges from 6.9% to 16.1%. This inconsistency may be because of the difference in the time series data that are modeled. But one can note that simple naïve models can act as more than just base case scenarios for short term forecasting of volatile time series such as truckload spot rates.

Table 9. Comparison of Best Fit Models Forecast and National Naïve Forecast

Region	**Best Fit Model of Each Region**	**MAPE for Best Fit Model of Each Region**	**MAPE for using National Naïve Forecast for All Regions**
National	FFNN	0.067	0.069
Chicago (604)	MA	0.111	0.109
Dallas (761)	ARIMA	0.154	0.161
Denver (804)	ARIMA	0.110	0.137
Los Angeles (917)	FFNN	0.093	0.098

6. Conclusions and Future Research

Forecasting short-term information for the truckload spot market has applications in many avenues. However, the volatile nature of the time series makes it difficult to predict. We model 4 time series forecasting techniques, namely Naïve, Moving Average, Auto-Regressive Moving Average, and Feed-Forward Neural Network to predict Spot Premium Ratio at the national level and New Premium Ratio for 4 high volume 3-Zip origin regions. Furthermore, we update the models regularly to make better predictions even when the market shifts. We contribute to the literature of truckload spot rate forecasting by creating weekly short term forecasts as opposed to monthly long term ones. Additionally, we contribute by investigating how updating the models, to account for changes in the underlying structure of the time series, affects the performance.

The results of the forecasting models indicate that there is no single model and combination of parameters (input variable, lags of input variable, training method, and updating frequency) that consistently produces minimum errors for all regions. Thus, we revert to the least complex forecasting method and use the naïve forecast of National SPR as the forecast for SPR/NPR values of all regions. This gives us performance that is comparable to the best fit models for each individual region. We also observe that using auto-lags of previous $1-4$ weeks as inputs to any model are sufficient. In practice, stakeholders do not look for the optimal solution, instead they want good enough models that are easy to implement. Complex models also have the hidden costs of less people who use the system understanding them and trusting the outputs. The parameters will be checked and updated less frequently, and the cost for making frequent forecasts will be higher. Thus, we can say that for making frequent short term forecasts of volatile time series data, simple models like Naïve models serve as more than just base cases. Moreover, in case of lack of other information auto-lags of recent past are suitable as inputs to any model.

Another key finding was the weak performance of neural networks under concept drift. When the models are trained on a period with stable values they produce outputs within that range of values even when the actual time series drastically increases or decreases. To make short term forecasts of volatile time series that undergo cyclical changes in the underlying data structure it is necessary to update our models to handle that concept drift. This can be extended to forecasting other real life volatile time series such as stock market prices, financial indicators, commodity prices etc.

There are many opportunities for further research. We describe 3 in some detail.

1. An extension to this research could be exploring other external input variables in the FFNN models. We have justified the use of recent auto-lags in our models. However, many researchers have successfully used a multitude of external variables like load type, market characteristics, day and month of shipment, origin and destination characteristics, economic indicators etc. to produce long term forecasts of spot rates. These variables may also prove useful for making short term forecasts. It can be worthwhile to check if they show any causation in the short term volatility of the spot market. And whether simple naïve forecasts are still sufficient.

2. The current research can also be strengthened by quantifying the benefits of the forecasting models in terms of savings in transportation costs. This can help identify if the naïve forecast of national SPR is a good enough estimate or whether more involved models like neural networks are justified for frequent short term forecasts of each region. This can be done for various contractual scenarios, like conventional contracts, index based contracts, future's market etc.

3. An interesting direction of research is to study spatio-temporal forecasting models that predict how the truckload spot rates change in the entire network with time. Such models can incorporate the effects of changes in one lane on its neighboring lanes. It can be a good tool to see epicenters of changes in the truckload spot market and how they travel through the network. This can be useful to identify lanes with higher spot rates. Carriers can use this information to capture demand in those lanes and shippers can identify where routing guides may fail and they may have to pay high spot rates.

Bibliography

Özkaya, E., Keskinocak, P., Joseph, R. V., & Weight, R. (2010). Estimating and Benchmarking Less-Than-Truckload Market Rates. *Transportation Research , Part E*(46), 667-682.

American Trucking Association. (n.d.). *trucking.org*. Retrieved April 29, 2019, from https://www.trucking.org/News_and_Information_Reports_Industry_Data.aspx

Asche, F., & Guttormsen, A. G. (2002). *Lead Lag Relationships Between Futures and Spot Prices.* Stiftelsen for Samfunns- og Naeringslivsforskning, Bergen (Norway). ETDE.

AT Kearney. (2018). *State of Logistics Report.* Council of Supply Chain Management Professionals.

Ballou, R. H. (1991). The Accuracy in Estimating Truck Class Rates for Logistical Planning. *Transportation Research, 25A*(6), 327-337.

Budak, A., Ustundag, A., & Guloglu, B. (2017). A Forecasting Approach for Truckload Spot Market Pricing. *Transportation Research, Part A*(97), 55-68.

Caplice, C. (2007). Electronic Markets for Truckload Transportation. *Production and Operations Management, 16*(4), 423-436.

Caplice, C., & Sheffi, Y. (2005). Combinatorial Auctions for Truckload Transportation. In P. Cramton, Y. Shoham, & R. Steinberg, *Combinatorial Auctions.* MIT Press Scholarship.

Cerqueira, V., Torgo, L., Oliveira, M., & Pfahringer, B. (2017). Dynamic and Heterogeneous Ensembles for Time Series Forecasting. *International Conference on Data Science and Advanced Analytics.* Tokyo: IEEE.

Corridore, J., & Chuah, S. Y. (2018). *Industry Surveys - Road & Rail.* Center for Financial Research & Analysis .

Costello, B. (2017). *Truck Driver Shortage Analysis.* American Trucking Associations.

Žliobaitė, I. (2010). *Learning Under Concept Drift: an Overview.* Computing Research Repository.

Gu, S., Tan, Y., & He, X. (2013). Recentness Bias Learning for Time Series Forecasting. *Information Sciences, 237*, 29-38.

Guo, W., Xu, T., Tang, K., Yu, J., & Chen, S. (2018). Online Sequential Extreme Learning Machine with Generalized Regularization and Adaptive Forgetting Factor for Time-Varying System Predictions. *Mathematical Problems in Engineering.*

Harding, M. (2017, June 29). *Drinking From The Transportation Market Fire Hose - Part 2.* Retrieved May 2019, from chainalytics.com: https://www.chainalytics.com/drinking-transportation-market-fire-hose-part-2/

HDT Staff. (2019, May). *Fleet Management*. Retrieved from truckinginfo.com: https://www.truckinginfo.com/328545/trucking-freight-futures-contracts-launched-on-a-commodities-exchange

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks, 2*, 359-366.

Kölmez, M. A., & Navruz, I. (2015). Forecasting the Day-Ahead Price in Electricity Balancing and Settlement Market of Turkey by Using Artificial Neural Networks . *Turkish Journal of Electrical Engineering & Computer Sciences, 23*, 841-852.

Kay, M. G., & Warsing, D. P. (2009). Estimating LTL rates using publicly available emperical data. *International Journal of Logistics: Research and Applications, 12*(3), 165-193.

Kiani, K. M., & Kastens, T. L. (2008). Testing Forecast Accuracy for Foreign Exchange Rates: Predictions from Feed Forward and Various Recurrent Neural Network Architectures. *Computational Economics, 32*, 383-406.

Klinkenberg, R. (2005). Meta-Learning, Model Selection, and Example Selection in Machine Learning Domains with Concept Drift. *LWA.* Saarbrücken.

Klinkenberg, R., & Renz, I. (1998). *Adaptive Information Filtering: Learning in the Presence of Concept Drift.* AAAI.

Kolter, J. Z., & Maloof, M. A. (2007). Dynamic Weighted Majority: An Ensemble Method for Drifting Concepts. *Journal of Machine Learning Research, 8*, 2755-2790.

Kristjanpoller, W., & Minutolo, M. C. (2015). Gold Price Volatility: A Forecasting Approach Using the Artificial Neural Network-GARCH Model. *Expert Systems with Applications, 42*, 7245-7251.

Li, J., & Parsons, M. G. (1997). Forecasting Tanker Freight Rates Using Neural Networks. *Maritime Policy & Management, 24*(1), 9-30.

Lindsey, C., Frei, A., Alibabai, H., Mahmassani, H. S., Park, Y.-W., Klabjan, D., . . . Keating, T. (2013). Modeling Carrier Truckload Freight Rates in Spot Market. Washington D.C.: Transportation Research Board.

Lyridis, D., Zacharioudakis, P., Mitrou, P., & Mylonas, A. (2004). Forecasting Tanker Market Using Artificial Neural Networks. *Maritime Economics & Logistics, 6*, 93-108.

Mendoza, A., & Ventura, J. A. (2009). Estimating Freight Rates in Inventory Replishment and Supplier Selection Decisions. *Logistics Research, 1*, 185-196.

Miller, J. W. (2018). ARIMA Time Series Models for Full Truckload Transportation Prices. *Forecasting, 1*, 121-134.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2016). *Time Series Analysis and Forecasting.* Wiley.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the Black Box: A Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks. *Ecological Modelling, 154*, 135-150.

Pickett, C. (2018). Navigating the US Truckload Capacity Cycle: Where are Freight Rates Headed and Why? *Journal of Supply Chain Management, Logistics and Procurement, 1*(1), 1-18.

Safi, S., & White, A. (2017). Short and Long-Term Forecasting Using Artificial Neural Networks for Stock Prices in Palestine: A Comparitive Study. *Electronic Journal of Applied Statistical Analysis, 10*(1), 14-28.

Sahin, B., Gürgen, S., Ünver, B., & Altin, I. (2018). Forecasting the Baltic Dry Index by Using an Artifical Neural Network Approach. *Turkish Journal of Electrical Engineering & Computer Sciences, 26*, 1673-1684.

Scott, A. (2015). The Value of Information Sharing for Truckload Shippers. *Transportation Research , 81*(Part E), 203-214.

Sheta, A. (2018). Prediction Stock Market Exchange Prices for the Reserve Bank of Australia Using Auto-Regressive with Exogenous Input Neural Network Model. *WSEAS Transactions on Systems, 17*, 2224-2678.

Swenseth, S. R., & Godfrey, M. R. (1996). Estimating Freight Rates for Logistics Decisions. *Journal of Business Logistics, 17*(1), 213-231.

Tsai, M.-T., Saphores, J.-D., & Regan, A. (2011). Valuation of Freight Transportation Contracts Under Uncertainty. *Transportation Research Part E, 47*, 920-932.

Tsymbal, A. (2004). *The Problem of Concept Drift: Definitions and Related Work.* Dublin: Alexey Tsymbal.

Uyar, K., Ilhan, Ü., & Ilhan, A. (2016). Long Term Dry Cargo Freight Rates Forecasting by Using Recurrent Fuzzy Neural Netwroks. *Procedia Computer Science, 102*, 642-647.

Widmer, G., & Kubat, M. (1996). Learning in Presence of Concept Drift and Hidden Contexts. *Machine Learning, 23*, 69-101.

Zeng, Q., Qu, C., Ng, A. K., & Zhao, X. (2016). A New Approach for Baltic Dry Index Forecasting Based on Emperical Mode Decomposition and Neural Networks. *Maritime Economics & Logistics, 18*, 192-210.

Zhang, Z., Ouyang, Z., Ma, C., Xu, J., Li, L., & Wen, Q. (2016). A Novel Combination Forecast Method on Non-Stationary Time Series and its Application to Exchange Rate Forecasting. *International Journal of Simulation - Systems, Science & Technology, 17*(47), 1.1-1.7.

Appendices

Appendix A – FFNN Architecture

Figure 11 shows the architecture of the neural network used in our analysis. We look at one example in depth here. We consider forecasting of national SPR, using auto-lags of previous 1 – 4 weeks as input. We look at forecasts made for one week out using the Single Training method. The values of the weights and biases are shown in Table 10. We calculate the relative importance of each input using methods described by Olden & Jackson (2002). The results in Table 11 show that all inputs have influence in the same order of magnitude, and that is the case for most other combination of parameters as well.

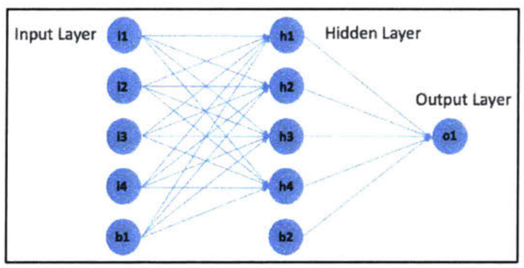


Figure 11. FFNN Architecture

Table 10. Weights and Biases of Connections Between (a)Input Layer and Hidden Layer (b)Hidden Layer and Output Layer

	i1	i2	i3	i4	b1
h1	0.7656	0.3157	1.1925	1.1973	-6.4468
h2	0.7632	-0.9097	1.4432	0.7955	-2.1309
h3	-0.9516	-0.4088	-0.8645	1.4302	-0.1021
h4	1.1059	0.0396	1.2387	-0.4521	2.8989

(a)

	h1	h2	h3	h4	b2
o1	-0.0288	0.8546	0.9138	-0.2630	-0.3277

(b)

Table 11. Relative Importance of Inputs

	i1	i2	i3	i4
Relative Importance	0.2665	0.1123	0.3464	0.2747

Appendix B – Best Fit Model for each Region and Forecast Horizon

Table 12. Best Fit Models for all Forecast Horizon for (a)National (b)Chicago (604) (c)Dallas (761) (d)Denver (804) (e)Los Angeles (917)

Forecast Horizon	Best Fit Model	MAPE 17	MAPE 18
1	Naïve	0.032	0.031
2	Naïve	0.047	0.049
3	Naïve	0.056	0.065
4	FFNN: input – SPR; input lag – 6; expanding window; update – 4 weeks	0.053	0.075
5	FFNN: input – SVR; input lag – 4; expanding window; update – 4 weeks	0.059	0.059
6	FFNN: input – SVR; input lag – 5; expanding window; update – 4 weeks	0.051	0.065
7	FFNN: input – SRI; input lag – 4; expanding window; update – 4 weeks	0.053	0.070
8	FFNN: input – CRI; input lag – 2; expanding window; update – 4 weeks	0.051	0.068

(a)

Forecast Horizon	Best Fit Model	MAPE 17	MAPE 18
1	FFNN: input – RRI; input lag – 6; expanding window; update – 12 weeks	0.084	0.084
2	FFNN: input – NRI; input lag – 2; expanding window; update – 4 weeks	0.099	0.097
3	FFNN: input – RRI; input lag – 4; expanding window; update – 4 weeks	0.109	0.085
4	FFNN: input – RRI; input lag – 3; expanding window; update – 8 weeks	0.096	0.088
5	FFNN: input – NVR; input lag – 2; expanding window; update – 8 weeks	0.105	0.087
6	FFNN: input – NVR; input lag – 5; expanding window; update – 8 weeks	0.094	0.090
7	FFNN: input – NVR; input lag – 1; expanding window; update – 4 weeks	0.107	0.090
8	FFNN: input – NPR; input lag – 5; expanding window; update – 12 weeks	0.114	0.089

(b)

Forecast Horizon	Best Fit Model	MAPE 17	MAPE 18
1	ARIMA (1, 0, 0): single training	0.161	0.131
2	ARIMA (1, 0, 0): single training	0.177	0.133
3	FFNN: input – NPR; input lag – 1; single training	0.165	0.133
4	ARIMA (1, 0, 0): single training	0.178	0.133
5	ARIMA (1, 0, 0): single training	0.178	0.133
6	ARIMA (1, 0, 0): single training	0.178	0.133
7	FFNN: input – NPR; input lag – 4; single training	0.166	0.143
8	ARIMA (1, 0, 0): single training	0.178	0.133

(c)

Forecast Horizon	Best Fit Model	MAPE 17	MAPE 18
1	ARIMA (1, 0, 0): expanding window; update – 12 weeks	0.093	0.107
2	ARIMA (1, 0, 0): expanding window; update – 4 weeks	0.101	0.114
3	FFNN: input – RRI; input lag – 5; single training	0.101	0.114
4	FFNN: input – NRI; input lag – 4; single training	0.097	0.124
5	ARIMA (1, 0, 0): expanding window; update – 4 weeks	0.103	0.120
6	FFNN: input – NRI; input lag – 1; single training	0.097	0.120
7	FFNN: input – NVR; input lag – 6; single training	0.100	0.122
8	FFNN: input – NVR; input lag – 3; expanding window; update – 12 weeks	0.102	0.109

(d)

Forecast Horizon	Best Fit Model	MAPE 17	MAPE 18
1	FFNN: input – NRI; input lag – 1; single training	0.076	0.075
2	FFNN: input – RRI; input lag – 6; expanding window; update – 4 weeks	0.081	0.089
3	FFNN: input – NRI; input lag – 1; single training	0.088	0.078
4	FFNN: input – NPR; input lag – 6; expanding window; update – 4 weeks	0.080	0.093
5	FFNN: input – NRI; input lag – 2; expanding window; update – 4 weeks	0.071	0.096
6	FFNN: input – NPR; input lag – 6; expanding window; update – 12 weeks	0.082	0.086
7	FFNN: input – NRI; input lag – 2; expanding window; update – 12 weeks	0.081	0.093
8	FFNN: input – RRI; input lag – 6; expanding window; update – 4 weeks	0.067	0.091

(e)

Appendix C – Comparison of Naïve Models

In Table 13 we tabulate the MAPE for all forecast horizons in the case when Naïve forecast of National SPR is used as the forecast for all regions; and the case where each region's SPR or NPR is predicted using its own Naïve model. We see that there are only minor differences in the two cases.

Table 13. Comparison of MAPEs of Naïve Forecasts

		National Naïve Model For All								Separate Regional Naïve Models							
		Prediction Number															
Region	MAPE	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
National	2017	0.03	0.05	0.06	0.06	0.07	0.07	0.08	0.08	0.03	0.05	0.06	0.06	0.07	0.07	0.08	0.08
	2018	0.03	0.05	0.06	0.08	0.09	0.09	0.10	0.11	0.03	0.05	0.06	0.08	0.09	0.09	0.10	0.11
Chicago (604)	2017	0.08	0.09	0.10	0.10	0.11	0.11	0.12	0.12	0.09	0.09	0.10	0.11	0.11	0.12	0.12	0.12
	2018	0.11	0.11	0.11	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.12
Dallas (761)	2017	0.17	0.17	0.16	0.17	0.17	0.18	0.19	0.20	0.16	0.18	0.18	0.18	0.18	0.18	0.18	0.18
	2018	0.15	0.15	0.15	0.14	0.14	0.15	0.15	0.16	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Denver (804)	2017	0.13	0.12	0.12	0.12	0.13	0.12	0.12	0.13	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	2018	0.13	0.15	0.15	0.16	0.16	0.15	0.15	0.15	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12
Los Angeles (917)	2017	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.06	0.08	0.08	0.08	0.09	0.09	0.09	0.09
	2018	0.09	0.10	0.10	0.11	0.12	0.12	0.12	0.13	0.09	0.11	0.10	0.10	0.12	0.11	0.11	0.11